ORSAY
N° d'ordre : 2541

THESE DE DOCTORAT D'ETAT

PRESENTEE

A L'UNIVERSITE PARIS XI
CENTRE D'ORSAY

POUR OBTENIR

LE GRADE DE DOCTEUR ES SCIENCES INFORMATIQUES

par

Fathi DEBILI

ANALYSE SYNTAXICO-SEMANTIQUE FONDEE SUR UNE ACQUISITION AUTOMATIQUE DE RELATIONS LEXICALES-SEMANTIQUES

Soutenue le 26 janvier 1982, devant la Commission d'Examen

MM.	P. DEBRAINE	Président
	A. ANDREEWSKY	Examinateurs
	G. GUIHO	
	D. KAYSER	Rapporteur
	L. MEHL	Examinateurs
	B. VAUQUOIS	

ORSAY
N° d'ordre : 2541

THESE DE DOCTORAT D'ETAT

PRESENTEE

A L'UNIVERSITE PARIS XI
CENTRE D'ORSAY

POUR OBTENIR

LE GRADE DE DOCTEUR ES SCIENCES INFORMATIQUES

par

Fathi DEBILI

ANALYSE SYNTAXICO-SEMANTIQUE FONDEE SUR UNE ACQUISITION AUTOMATIQUE DE RELATIONS LEXICALES-SEMANTIQUES

Soutenue le 26 janvier 1982, devant la Commission d'Examen

MM.	P. DEBRAINE	Président
	A. ANDREEWSKY	Examinateurs
	G. GUIHO	
	D. KAYSER	Rapporteur
	L. MEHL	Examinateurs
	B. VAUQUOIS	

UNIV.
LEIDEN
BIBL.

REMERCIEMENTS

Dans une formule simple mais profonde je souhaite présenter ici mes remerciements à tous ceux qui m'ont permis d'accomplir ce travail et dont il n'est pas possible, en peu de lignes, de décrire tout ce qu'ils m'ont apporté d'aide, de soutien et de gentillesse. Aussi, qu'il me soit permis seulement de les citer.

Aux membres du jury :

Monsieur Pierre DEBRAINE, Professeur à l'Institut National des Sciences et Techniques Nucléaires (INSTN), Professeur associé à l'Université Paris XI, Centre d'Orsay ;

Monsieur Alexandre ANDREEWSKY, Ingénieur au LIMSI (CNRS) ;

Monsieur Gérard GUIHO, Professeur à l'Université Paris XI, Centre d'Orsay ;

Monsieur Daniel KAYSER, Professeur à l'Université Paris XI, Centre d'Orsay ;

Monsieur Lucien MEHL, Conseiller d'Etat Honoraire, Président du Centre d'Informatique Juridique (CEDIJ), Directeur de l'ERA n°430 du CNRS ;

Monsieur B. VAUQUOIS, Professeur à l'Université Scientifique et Médicale de Grenoble (USMG), Directeur du Groupe d'Etude pour la Traduction Automatique (GETA) ;

Aux membres de l'équipe à laquelle j'appartiens :

Messieurs Christian FLUHR, Professeur associé à l'Université Paris XI, Centre d'Orsay ; Jean-Paul BINQUET et Bertrand POUDEROUX ;

A Monsieur Jean-Baptiste BERTHELIN, attaché de recherche au CNRS ;

A Monsieur Eugène CHARPENTIER, Adjoint au Directeur de l'INSTN ;

j'exprime encore ma profonde et sincère reconnaissance.

A Saclay le 17 Novembre 1981

Fathi DEBILI

PREAMBULE

Le présent travail s'inscrit dans l'ensemble des travaux d'une équipe constituée autour de Monsieur A. ANDREEWSKY et qui comprend MM. J.P. BINQUET C. FLUHR, B. POUDEROUX et moi-même. Elle bénéficie par ailleurs de l'appui bienveillant de M. P. DEBRAINE Professeur à l'Institut National des Sciences et Techniques Nucléaires, Professeur associé à l'Université de Paris XI et de M. L. MEHL Conseiller d'Etat Honoraire, Président du Centre d'Informatique Juridique, Directeur de l'ERA n° 430, formation dont je fais partie.

Cette équipe, dont les travaux se situent dans le cadre du traitement automatique des langues naturelles, a notamment orienté d'emblée ses applications vers la réalisation de systèmes documentaires.

En particulier l'idée d'une acquisition automatique des informations linguistiques a toujours guidé nos recherches. A cet égard, la démarche suivie fut de partir des niveaux les plus proches du texte pour aller vers des niveaux plus abstraits. C'est pourquoi nous nous posons maintenant le problème de l'acquisition des informations sémantiques après avoir abordé les niveaux plus spécialement syntaxique et morphologique.

ANALYSE SYNTAXICO-SEMANTIQUE FONDEE SUR UNE ACQUISITION AUTOMATIQUE DE RELATIONS LEXICALES-SEMANTIQUES

RESUME

Le travail présenté s'inscrit dans le cadre général du traitement automatique des langues naturelles dont les applications sont fort nombreuses et bien connues. C'est cependant la documentation automatique que nous avons plus particulièrement en vue.

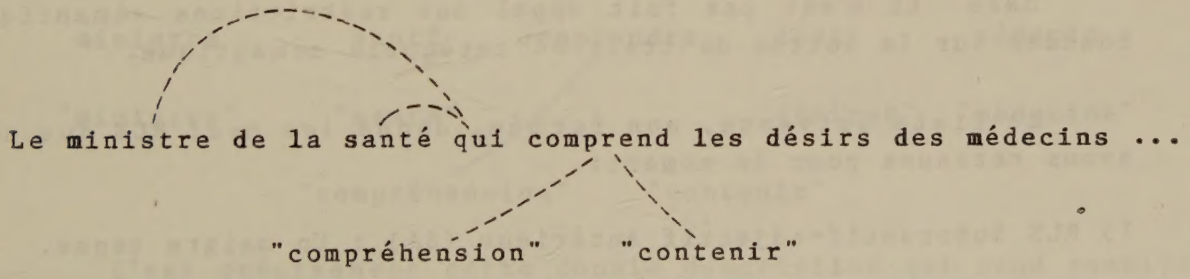
Dans toutes ces applications où les "matières brutes" sont des données textuelles, l'un des buts fondamentaux du traitement linguistique est d'aboutir à une ou plusieurs représentations de leur contenu qui soient exploitables, c'est à dire sur lesquelles on puisse faire un certain nombre d'opérations permettant d'atteindre l'objectif visé.

Dès lors, plusieurs problèmes liés se posent.

Le premier est relatif à la définition même de la représentation à laquelle nous souhaitons aboutir. Pour celà nous avons choisis délibérément une représentation dépendancielle directement inspirée des "stemmas" de L. TESNIERE.

Viennent ensuite les problèmes liés à la détermination automatique de cette représentation, c'est à dire à la reconnaissance automatique des diverses connexions qui s'établissent entre les éléments de la phrase.

On est alors confronté d'une part aux ambiguïtés lexicales et structurales, par exemple:



d'autre part aux difficultés liées à la définition, acquisition et utilisation des informations nécessaires pour résoudre ces ambiguïtés, et, partant, pour atteindre la représentation souhaitée.

Ce dernier point revêt un caractère fondamental dès lors que nous nous assignons en outre de pouvoir appréhender des univers textuels ouverts, c'est à dire sans aucune contrainte quant à leur nature ou à leur diversité, et ce, tant dans le vocabulaire que dans les constructions syntaxiques.

Un dernier aspect enfin, lié dans notre repère à l'utilisation des informations à mettre en oeuvre pour la détermination des connexions entre mots, s'impose : la détection des synonymies lexico-syntaxiques ou paraphrases. En effet, outre l'intérêt que cette détection présente en soi, elle est nécessaire si l'on veut conférer certaines capacités inférentielles aux informations qui pourraient être acquises. La représentation dépendancielle visée devra donc également permettre cette détection.

Ce sont là les points essentiels que nous abordons.

Après avoir survolé sous l'angle sémantique, dans le premier chapitre, les principales méthodes qui ont été généralement proposées pour résoudre les ambiguïtés lexicales et structurales, nous optons, dans le deuxième chapitre, pour une analyse fondée sur l'acquisition et l'utilisation de relations lexicales-sémantiques (RLS). Celles-ci correspondent aux relations qu'entretiennent entre elles les unités lexicales selon l'axe paradigmatique (ce sont alors des relations de parenté sémantique : synonymie, antonymie, etc.) ou syntagmatique (et ce sont alors des relations de compatibilité sémantique où l'expression idiomatique du sens est maintenue ; par exemple : **souhaiter une bonne année, désirer une jolie maison ; boucherie chevaline, sport équestre).**

Mais il n'est pas fait appel aux restrictions sémantiques fondées sur la notion de trait ou catégorie sémantique.

La liste suivante, non fermée, donne les neuf **RLS** que nous avons retenues pour le moment:

1) RLS Substantif-adjectif Antérieur (SA) : Un maigre repas.

2) RLS Substantif-adjectif Postérieur (SP): Un repas maigre.

3) RLS Substantif-Substantif (SS_p) : La porte de la maison.

4) RLS Verbe-Sujet (VS) : La turbine entraîne

5) RLS Verbe-complément Direct (VD): entraîne une pompe

6) RLS Verbe-complément Indirect (VI_p): rayé de la liste

7) RLS Verbe-Adverbe (VA) : transporter bénévolement

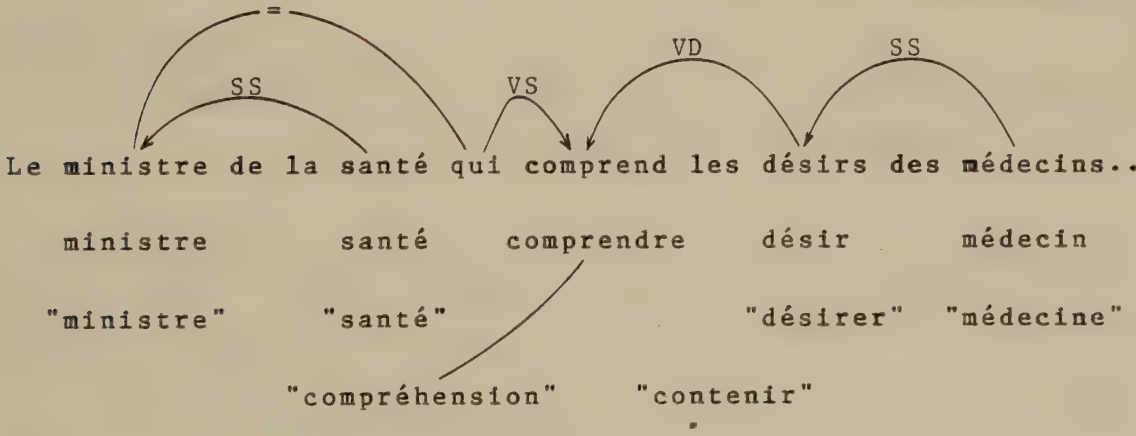
8) RLS Adjectif-Adverbe (AA) : férocement violent

9) RLS Adjectif-Complément d'adjectif (AC_p) : dédaigneux du danger

Les RLS paradigmatiques sont par contre vues sous l'angle de l'appartenance de plusieurs éléments à un même ensemble sémantique correspondant à la notion traditionnelle de famille de mots (par exemple : [ART, ARTISTE, ARTISTIQUE, ARTISTIQUEMENT]), éventuellement étendue, c'est à dire incluant également des éléments n'ayant pas même radical (par exemple : [ENTREE, INPUT]).

Décrite dans les termes de ces relations, la représentation d'une phrase se fait donc selon deux axes : paradigmatique et syntagmatique.

Ainsi, la phrase exemple citée plus haut recevra la description suivante (pour ne pas alourdir la figure, nous n'indiquons pas les relations "satellites" relatives aux mots vides) :



C'est précisément cette double description qui rend possible la détection des synonymies lexico-syntaxiques prises au sens large du terme (cf. chapitre 2 & 5).

Pour atteindre cette représentation, l'analyseur syntaxico-sémantique (ASS) proposé repose sur trois composantes :

- MORPHOLOGIQUE : Elle permet d'obtenir, à partir d'un lexique de mots, les familles de mots et, par suite, les descriptions paradigmatiques potentielles de la phrase.

Cette composante, décrite dans le chapitre 4, repose sur un découpage des mots en **préfixe + radical + suffixe.** La validation de ces découpages repose elle-même sur la notion de schéma dérivationnel (ensemble de suffixes ou de préfixes qui peuvent être associés à un radical commun) et sur l'emploi de règles de compatibilité ou d'incompatibilité des suffixes ou des préfixes. Par exemple : ATION et ETTE sont des suffixes incompatibles, c'est à dire que l'on ne peut pas trouver un radical qui puisse former avec eux, deux mots pouvant appartenir à un même ensemble sémantique.

- SYNTAXIQUE : Elle permet d'obtenir les différentes descriptions syntagmatiques potentielles de la phrase.

Décrite dans le chapitre 3, cette composante construit toutes les relations de dépendance potentielles internes à la phrase. Elle opère en trois étapes : dans la première, elle sélectionne les mots susceptibles de participer à une ou plusieurs relation(s). Chaque mot est alors supposé être le "dépendant" d'une ou plusieurs relations ; dans la deuxième, elle délimite le champ d'investigation ; dans la troisième, elle établit les relations potentielles, en sélectionnant les différents "gouverneurs" possibles.

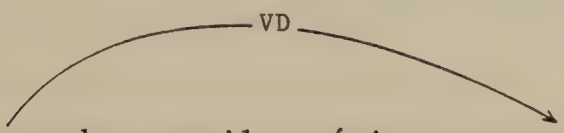
Une étape préalable consiste à segmenter chaque phrase en chaînes nominales et verbales d'une part, et à établir les relations "satellites" d'autre part.

- SEMANTIQUE : Décrite dans le chapitre 5, elle permet, d'une part, en relation avec la composante syntaxique, d'acquérir les RLS syntagmatiques, d'autre part, en faisant appel à un processus de raisonnement par analogie, de résoudre les ambiguïtés lexicales et structurales mises à jour respectivement par les deux composantes précédentes.

En effet, dans de très nombreux cas, la composante syntaxique seule suffit pour établir sans ambiguïté les relations de dépendance. Dès lors, l'acquisition des RLS syntagmatiques qui leur correspondent rend possible la résolution des ambiguïtés

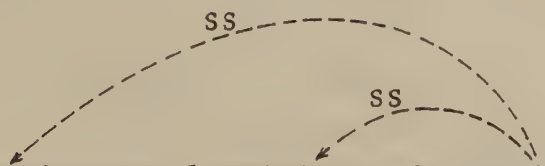
lexicales ou structurales qui peuvent par ailleurs se présenter.

Par exemple, les RLS qui peuvent être acquises à partir de la phrase:



Les membres du conseil supérieur sont nommés par le ministre.

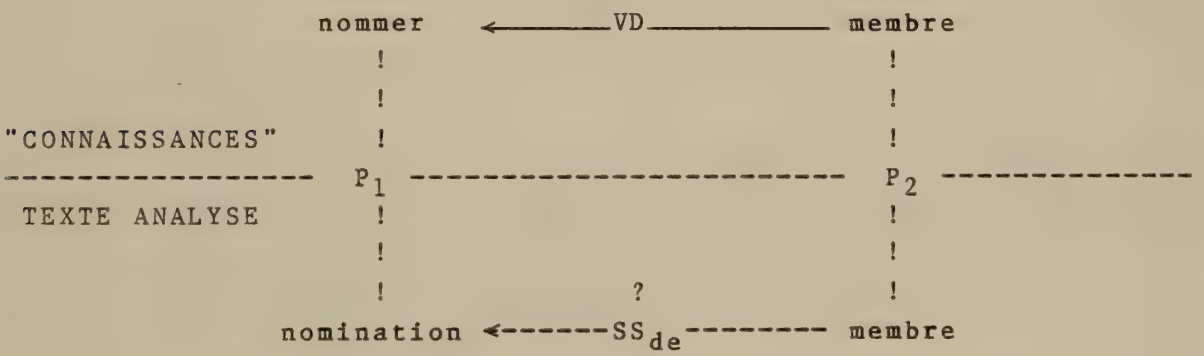
peuvent servir à lever l'ambiguïté structurale qui figure dans le fragment de phrase:



La nomination par le ministre des membres du conseil supérieur...

Ce processus met en oeuvre un mécanisme de raisonnement par analogie qui permet de simuler certaines règles transformationnelles (ici, la RLS (nommer, membre, VD), issue de la première phrase, permet de "préférer",dans la deuxième phrase, la dépendance candidate nomination des membres).

Ce mécanisme peut être représenté par le schéma suivant:



où P_1 et P_2 sont des relations de parenté sémantique (axe paradigmatique), VD une relation valide issue de la "connaissance", et SS la relation à valider.

C'est ce même mécanisme qui permet de résoudre aussi les ambiguïtés lexicales ou polysémies.

L'analyseur syntaxico-sémantique présenté se caractérise donc par une acquisition de l'information linguistique que l'on peut rattacher aux trois niveaux classiques:

- morphologique: par la génération automatisée de familles de mots ;
- syntaxique: par la simulation de règles transformationelles au moyen d'un raisonnement par analogie rendant possible la détection de certaines synonymies lexico-syntaxiques ;
- sémantique: par l'apprentissage de RLS syntagmatiques.

Dans la conclusion enfin, après avoir discuté des limites des outils que nous avons construits, deux aspects sont envisagés: - le premier est relatif aux applications possibles. Deux domaines sont en particulier abordés: la documentation automatique et la reconnaissance automatique de la parole ; - le deuxième est lié au caractère évolutif du système proposé et à l'acquisition éventuelle de nouvelles informations linguistiques: les classes ou catégories sémantiques d'une part, les mots composés pris comme un tout d'autre part.

TABLE DES MATIERES

Pages

INTRODUCTION 9

Chapitre premier: LE PROBLEME - LES SOLUTIONS ENVISAGEES

I. Reconnaissance des relations internes aux chaînes nominales 15
II. Reconnaissance de le relation "Sujet - Verbe" 20
III. Reconnaissance de la relation "Verbe - Complément". 22
IV. Reconnaissance des référents 23
V. Reconnaissance de la polysémie 26
VI. Conclusion 27

Chapitre deuxième: NOTIONS ET HYPOTHESES DE BASE

I. Notion de relation lexicale sémantique (RLS)..... 31
II. RLS paradigmatiques - RLS syntagmatiques..... 35
III RLS syntagmatiques..... 36
III.1. Introduction..... 36
III.2. Description des RLS syntagmatiques..... 38
III.2.1. RLS Substantif - adjectif Antérieur (SA)..... 39
III.2.2. RLS Substantif - adjectif Postérieur (SP)..... 41
III.2.3. RLS Substantif - Substantif (SS_p)..... 42
III.2.4. RLS Verbe - Sujet (VS)..... 45
III.2.5. RLS Verbe - complément Direct (VD)..... 47
III.2.6. RLS Verbe - complément Indirect (VI_p)..... 48
III.2.7. RLS Verbe - Adverbe (VA)..... 49
III.2.8. RLS Adjectif - Adverbe (AA)..... 50
III.2.9. RLS Adjectif - Complément d'adjectif (AC_p)..... 51
IV. Synonymie lexico-syntaxique..... 52
V. Conclusion..... 56

Pages

Chapitre troisième

COMPOSANTE SYNTAXIQUE DU SYSTEME

DESCRIPTION SYNTAGMATIQUE DE LA PHRASE

Introduction..... 58

Première Partie: SEGMENTATION DE LA PHRASE

I. Notion de filtre..... 63
I.1. Filtre élémentaire..... 63
I.2. Filtrage..... 65
I.3. Filtre syntaxique..... 66
I.4. Filtrage syntaxique..... 67
I.5. Extension de la notion de filtre..... 67
II. Resserrement du champ d'investigation..... 68
III. Segmentation de la phrase..... 70
III.1. Reconnaissance des chaînes nominales..... 72
III.1.1. Caractéristiques des chaînes nominales..... 72
III.1.2. Matrice de reconnaissance des chaînes nominales.. 72
III.1.2.1. Lignes et colonnes de la matrice..... 73
III.1.2.2. Corps de la matrice..... 74
III.1.3. Expérimentation..... 81
III.2. Reconnaissance des chaînes verbales..... 85

Deuxième Partie: CONSTRUCTION DES RELATIONS DE DEPENDANCE POTENTIELLES

IV. Analyse sous forme de dépendances..... 93
IV.0. Bruit - Silence..... 94
IV.1. Relations satellites..... 96
IV.2. Reconnaissance des relations de dépendance..... 99
IV.2.1. Analyse de la chaîne nominale..... 99
IV.2.1.1. Construction des relations satellites..... 100
IV.2.1.2. Construction des relations homosyntagmatiques potentielles..... 103

a) RLS Substantif - adjectif Antérieur (SA)..... 103
b) RLS Substantif - adjectif Postérieur (SP)..... 107
c) RLS Substantif - Substantif (SS_p)..... 111
d) RLS Verbe - Sujet (VS); Verbe - compl. Direct (VD) Verbe - Complément Indirect (VI_p)..... 112
e) RLS Verbe - Adverbe (VA) & Adjectif - Adverbe (AA) 117
f) RLS Adjectif - Complément d'adjectif (AC_p)..... 117

IV.2.1.3. Eléments libres..... 121

IV.2.2. Analyse de la chaîne verbale..... 123

IV.2.3. Construction des relations hétérosyntagmatiques potentielles..... 131

IV.2.3.1. Reconnaissance des relations VS potentielles..... 133

a) dans les formes actives
1°) Reconnaissance dans les formes MP..... 135
2°) Reconnaissance dans les formes MIP..... 140
3°) Reconnaissance dans les formes MII..... 140

b) dans les formes passives..... 143

IV.2.3.2. Reconnaissance des relations VD potentielles..... 145

a) dans les formes actives..... 145

b) dans les formes passives..... 151

IV.2.3.3. Reconnaissance des relations VI_p potentielles.... 151

IV.2.4. Analyse de l'anaphorique "qui"..... 154

IV.2.4.1. Reconnaissance de l'antécédent..... 154

IV.2.4.2. Les options..... 156

IV.2.4.3. Résultats de l'expérimentation..... 156

IV.2.5. Enchaînement des procédures..... 161

V. Conclusion..... 161

Chapitre quatrième

COMPOSANTE MORPHOLOGIQUE DU SYSTEME

CONSTRUCTION AUTOMATIQUE DE FAMILLES DE MOTS

DESCRIPTION PARADIGMATIQUE DE LA PHRASE

I. Introduction..... 174

II. Notion de familles de mots..... 176

III. Principe de la méthode proposée pour l'élaboration automatique des familles de mots..... 177

IV. Reconnaissance automatique de familles de mots..... 177

IV.1. Familles de mots primitives..... 178
IV.2. Choix du découpage d'un mot en radical + suffixe... 184
IV.3. Obtention des familles de mots propres..... 186
IV.3.1. Notion de schéma dérivationnel..... 188
IV.3.2. Explosion des familles de mots primitives..... 189
IV.3.2.1. Schéma dérivationnel impropre..... 191
Hypothèse I 192
IV.3.2.2. Matrice de corrélation des suffixes..... 192
IV.3.2.3. Matrice d'incompatibilité des suffixes - Matrice de compatibilité des suffixes..... 193
IV.3.2.4. Production des familles de mots propres..... 197
Hypothèse II 199
IV.3.3. Obtention des familles de mots suffixales complètes 203
IV.3.3.1. Notion de polysémie..... 205
IV.3.3.2. Réunion transitive..... 205
IV.3.3.3. Choix du radical..... 206
IV.3.3.4. Notion de suffixes composés..... 208
IV.3.3.5. Choix du représentant..... 209
V. Traitement des préfixes..... 212
VI. Obtention des familles de mots complètes..... 212
VII. Gestion des familles de mots..... 215
VIII. Bilan de l'expérience..... 215
IX. Conclusion..... 219
X. Description paradigmatique..... 219

Chapitre cinquième

COMPOSANTE SEMANTIQUE DU SYSTEME

ACQUISITION ET UTILISATION DES RELATIONS LEXICALES SEMANTIQUES

I. Acquisition automatique des RLS syntagmatiques..... 226
I.1. Hypothèse de base..... 226
I.2. Apprentissage des RLS syntagmatiques..... 228
I.3. RLS induites par la syntaxe..... 229
II. Représentation des connaissances..... 232

III. Mécanisme de résolution: raisonnement par analogie. 236
IV. Résolution des ambiguïtés syntaxiques..... 237
IV.1. Mécanisme de résolution 237
IV.2. Evaluation de la note de validité..... 239
IV.3. Apprentissage - Organigramme général du traitement syntaxico - sémantique..... 242
IV.4. Expérimentation..... 244
IV.4.1. Exemples de résolution..... 244
IV.4.2. Performances du traitement sémantique..... 244
IV.4.3. Evolution de l'apprentissage sémantique..... 249
V. Résolution de la polysémie..... 250
V.1. Mécanisme de résolution..... 252
V.2. Evaluation de la note..... 253
V.3. Expérimentation..... 254
VI. Conclusion..... 255

CONCLUSION GENERALE

I. Les limitations..... 259
II. Les applications..... 261
III. Les espoirs..... 264
a) sur l'acquisition automatique de nouvelles informations linguistiques..... 264
b) sur l'utilisation de cette nouvelle connaissance... 267
IV. Caractère évolutif du système..... 270

ANNEXES

LEXIQUE DES FIGURES 272
LEXIQUE TERMINOLOGIQUE 277
CODAGE DES LETTRES ACCENTUEES 278

BIBLIOGRAPHIE 279

INTRODUCTION

"..., comprendre une phrase, c'est saisir l'ensemble des connexions qui en unissent les différents mots."

Lucien Tesnière.

Par le choix de cette phrase introductive, nous voulons, non pas, bien sûr, marquer l'aspect suffisant des connexions, ni donc adhérer à la définition de la compréhension qui est ainsi proposée, mais plutôt souligner la nécessité qu'il y a d'établir ces connexions si l'on veut emprunter les "méandres" de la compréhension ou s'y engager davantage.

En ce sens, les travaux que nous allons décrire ambitionnent seulement d'apporter une contribution à l'effort général qui tend à expliquer ou simuler le phénomène de la compréhension. Plus modestement, mais plus justement , il s'agit de construire un outil qui, penson-nous, serait utile pour la simulation sur machine de ce phénomène. Aussi pouvons-nous préciser, dès maintenant, que nous n'aurons pas à revenir, dans la suite, sur cette notion de compréhension ; mais que bien évidemment elle restera, comme pour beaucoup d'autres recherches, le but à atteindre, plus ou moins lointain d'ailleurs, selon le niveau requis par les applications envisagées.

Par cet exergue nous voulons également circonscrire d'emblée, dans le cadre général que représente le traitement automatique des langues naturelles, les préoccupations qui seront les nôtres dans les pages qui suivent, et qui ressortissent précisement à la reconnaissance automatique des connexions qui s'établissent entre les mots de la chaîne écrite.

En effet, dans les applications à caractère linguistique orientées vers l'appréhension de données textuelles, l'un des buts fondamentaux du traitement linguistique est d'aboutir à une ou plusieurs représentations de leur contenu qui soient exploitables, c'est-à-dire sur lesquelles on puisse faire un certain nombre d'opérations permettant d'atteindre l'objectif fixé. Dans cette représentation, les mots ainsi que les relations qui les unissent sont généralement un élément essentiel.

De nombreuses recherches se sont engagées dans cette voie depuis fort longtemps (*). Nous relaterons dans le chapitre I les principales méthodes qui furent proposées pour venir à bout de certaines difficultés rencontrées sur ce chemin.

(*) Pour un rappel "historique" voir notamment (B. VAUQUOIS, 109)

Parmi ces difficultés, l'une des principales est liée au volume considérables informations linguistiques qu'il y a lieu de définir et d'enregister pour pouvoir effectuer automatiquement la reconnaissance de ces relations.

Pourtant, on peut avancer dès maintenant (et sans que cela soit une critique, car c'est dans ce sens qu'il fallait opérer) que, dans la plupart des expériences menées, l'on s'est davantage préoccupé de la nature des informations nécessaires à cette reconnaissance et de leur représentation et utilisation, plutôt que de leur acquisition. L'attitude qui prévalut fut donc, le plus souvent, de recourir à des méthodes d'acquisition manuelles, ceci ayant dès lors pour conséquence de limiter fortement les univers textuels sur lesquels devaient porter les expériences.

Or, si cette limitation portant sur la nature ou la diversité des données textuelles traitées peut, pour certaines applications, être sans risque, voire recherchée (nous pensons ici à des applications ponctuelles ; un robot ne doit pas être forcément polyvalent, par exemple), elle met en cause par contre d'autres applications qui, de par leur nature même, sont appelées à appréhender des univers ouverts. C'est par exemple notamment le cas des applications de traduction ou de documentation automatiques en général.

Nous voulons ainsi souligner un deuxième aspect important de nos préoccupations : à savoir l'élaboration de méthodes d'acquisition automatique des informations dont nous aurons besoin pour établir les connexions(*), ce qui est notre objectif principal.

Cependant, bien que les outils que nous souhaitons développer puissent être orientés vers toutes les applications mettant en oeuvre le langage naturel c'est plus particulièrement le domaine de la documentation automatique que, pour des raisons "historiques" et pratiques, nous aurons en vue.

A cet égard, nous pourrions relever que, dans ce domaine, plus qu'ailleurs, ce sont les connexions, c'est à dire les relations syntagmatiques, qui ont été le plus souvent sacrifiées.

La représentation des données textuelles que nous tenterons d'obtenir sera donc supportée, notamment, par une description syntagmatique dépendancielle. Elle consistera en un graphe dont les noeuds seront les mots pleins et les arcs les relations ou connexions qui les unissent.

Pour définir cette représentation, nous nous inspirerons délibérément de la grammaire traditionnelle et, notamment, des travaux de L. Tesnière.

(*) mais aussi (voir la dernière partie de la conclusion) pour assurer au système une capacité évolutive dans les différents aspects qui sont les siens.

Comme nous le verrons dans le chapitre V, cette représentation devra permettre en particulier la simulation de certaines règles transformationnelles. La détection automatique de certaines paraphrases sera alors possible. Outre l'intérêt que celle-ci représente en soi, nous verrons aussi que c'est par cette voie que nous pourrons attribuer quelques capacités inférentielles aux informations qui seront acquises.

Notre représentation devra donc inclure en outre une description paradigmatique. Dans le chapitre IV nous en donnerons la nature.

Auparavant, nous présenterons dans le chapitre II ces informations qu'il est utile d'acquérir pour construire cette représentation. Concernant leur nature, nous nous inspirerons des "fonctions lexicales" de A.K. ZHOLKOVSKIJ et I.A. MEL'CHUK. En particulier, nous ne ferons pas appel à des catégories ou traits sémantiques : pas parce que celles-ci seraient inutiles, mais plutôt en raison de la double difficulté qu'il y a à les définir et à les rassembler. Nous verrons d'ailleurs dans la conclusion que les matériaux que nous aurons acquis automatiquement pourront permettre éventuellement la détection de telles classes

Les chapitres III et IV seront respectivement consacrés à la présentation de ce que nous appellerons les composantes syntaxique et morphologique du système. La première composante devra fournir la ou les descriptions syntagmatiques potentielles, la seconde la ou les descriptions paradigmatiques potentielles.

C'est enfin la composante sémantique du système, décrite dans le chapitre V, qui permettra, moyennant les informations qui auront été acquises entre temps, de dégager les représentations à retenir.

Peut-être est-il enfin important de rappeler que, dès lors que nous voulons appréhender des univers textuels ouverts, les algorithmes que nous chercherons à élaborer ne pourront en aucune manière fournir des résultats avec un taux d'efficacité égal à 100 % . Autrement dit, nous devrons nous accommoder d'un certain taux d'erreur ; mais celui-ci ne devra évidemment pas dépasser un certain seuil tenu pour acceptable pour les applications envisagées.

Ajoutons que les traitements que nous allons décrire se situent à la suite des traitements morpho-grammaticaux qui ont été très largement décrits dans les thèses ou publications suivantes (2, 3, 5, 46, 60, 68) (cf. chapitre III, Introduction). En outre, la notion de filtre introduite dans (4, 6(a), 6(b)) peut être considérée comme l'une des hypothèses linguistiques de base du système présenté.

Chapitre premier

LE PROBLEME

LES SOLUTIONS ENVISAGEES

Dans ce chapitre on tente de définir brièvement les problèmes que l'on se pose. On essaye en même temps de donner les solutions qui sont généralement proposées pour résoudre ces problèmes, et ce de façon ponctuelle et au travers d'exemples concrets. Nous ne rappellons pas cependant les différentes expériences qui ont été menées ici ou là et qui avaient pour objet le traitement du langage ou d'un sous-ensemble de celui-ci. Elles sont bien connues et un rappel en est fait dans la thèse de J.H. JAYEZ (76 pages en français) **(70)**

I. Reconnaissance des relations internes aux chaînes nominales•(*):

Considérons le segment de phrase suivant :

(1) *La vente des produits de beauté aux clients étrangers*

A ce segment de phrase, nous pouvons associer la structure syntaxique superficielle suivante :

ARTD S PREPART S PREPART S PREPART S ADJP

où :

ARTD veut dire ARTicle Défini ;
S Substantif ;
PREPART PREPosition ARTicle ;
ADJP ADJectif Postérieur.

Ceci suppose que nous ayons levé l'ambiguïté grammaticale (cf. chap. I Introduction) sur certains mots tels que "la" qui peut par ailleurs être pronom préverbal ou même substantif.

Le problème que l'on se pose est celui de la reconnaissance des relations, dites de dépendance, internes à la chaîne nominale et que l'on appelle quelquefois parenthésage ou fléchage.

Ainsi le parenthésage de (1) donne :

(1a) (La vente (des produits de beauté) (aux clients étrangers))

Le niveau d'emboitement des parenthèses permet d'indiquer l'ordre selon lequel sont effectués les regroupements.

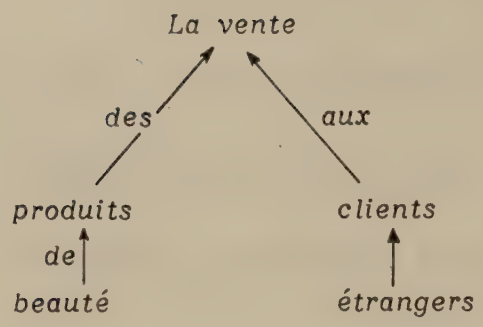
Ici nous aurons :

des produits de beauté
aux clients étrangers
la vente des produits de beauté
la vente aux clients étrangers

Notons que l'on peut trouver une représentation fléchée ou arborescente équivalente à la représentation parenthésée (B. VAUQUOIS, 109, p.55). Nous aurions ainsi, pour (1), la représentation arborescente suivante :

(*) Toutes les notions ou notations surmontées d'un point noir sont définies ou reprises dans le lexique donné en annexe.

(1b)



la représentation fléchée étant :

(1c) La vente des produits de beauté aux clients étrangers

Le problème réside précisément dans l'obtention de l'une quelconque de ces représentations.

En effet, si l'on considère le segment de phrase suivant :

(2) La vente des produits de beauté aux algues marines
ARTD S PREPART S PREPART S PREPART S ADJP

nous nous rendons compte qu'il lui correspond une représentation fléchée différente (cf. (2a)) bien que la structure syntaxique superficielle qui lui est associée soit identique à celle de (1).

(2a) La vente des produits de beauté aux algues marines

Ainsi, l'usage seul de critères syntaxiques ne suffit-il pas pour obtenir de manière non-ambiguë ces diverses relations.

On admet alors communément qu'il y a lieu d'introduire une composante sémantique.

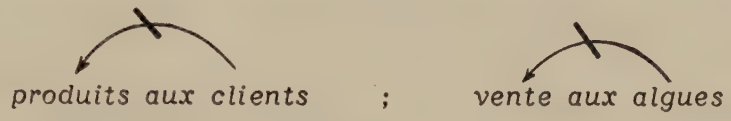
Celle-ci consiste en général en une contrainte supplémentaire que l'on doit respecter lors de la mise en relation des mots de la phrase.

Ainsi, si l'on veut éviter d'établir des relations parasites :

dans (1) et

dans (2)

on peut par exemple introduire les contraintes d'incompatibilité suivantes :



qui signifient simplement que de telles relations sont fausses et que par conséquent elles ne peuvent exister.

En fait les contraintes sémantiques utilisées reposent traditionnellement sur ce qu'il est convenu d'appeler traits ou marqueurs sémantiques (KATZ & FODOR, 71)

Ces traits sémantiques, utilisant la correspondance qu'il y a entre les mots et les signifiés (PITRAT, 95), permettent de spécifier la nature des objets visés.

Un mot peut recevoir un ou plusieurs traits sémantiques.

Les traits *(humain)*, *(non-humain)*, *(animé)*, *(non-animé)*, *(comestible)*, etc. sont bien connus.

Il est important de noter cependant la difficulté qu'il y a d'établir ces marqueurs sémantiques.

Si certains traits sémantiques, tels ceux mentionnés plus haut, semblent être admis de tous, il n'en est malheureusement pas toujours de même pour tous les traits définis. Ainsi en témoigne le trait **(animal)** qui, sans doute, pourrait être attribué à l' **"homme"** sans choquer certains, mais que d'autres ne pourraient accepter.

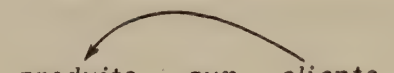
Souvent, en effet, on remarque que les traits sémantiques proposés sont induits par les problèmes à résoudre. Ils perdent alors toute chance de présenter un caractère universel.

Il n'est pas rare non plus de constater que plusieurs jeux de traits sémantiques peuvent être proposés pour résoudre un seul et même problème. Ceci laisse donc apparaître une certaine subjectivité qui semble, du moins pour l'instant, être inhérente à la définition des traits sémantiques.

Un autre aspect de la difficulté réside dans la convergence de ce processus de définition : les traits obtenus doivent en effet présenter un certain caractère de stabilité et leur nombre doit être en tout état de cause relativement petit.

L'utilisation qui est faite des traits sémantiques pour éviter les relations parasites, consiste en général en une vérification de la compatibilité ou non des traits associés aux mots mis en relation.

Pour éliminer la relation :



dans (1)

nous pourrions par exemple faire appel à l'organisation du vocabulaire que propose G. SABAH, et qui consiste, en l'occurrence, à attribuer à **"produits"** certaines "qualités" parmi lesquelles on ne trouve pas :

"matière"
produit ---------- *humain*

"humain" étant une qualité attribuée à **"clients"** .

Ceci nous montre par ailleurs qu'il y a lieu de différencier la nature des liens entre les unités lexicales et les traits associés (G. SABAH, 101).
L'interprétation de "de" ne conduit pas toujours à une relation de possession : **"Table de bois"** . Pour conférer à cette organisation du vocabulaire un pouvoir inférentiel, G. SABAH introduit en outre une "loi de composition interne" opérant sur ces liens ; par exemple :

possession + utilité = capacité

Ainsi : Un animal possède des yeux
les yeux servent à voir
donc l'animal peut voir.

Bien évidemment, comme presque toute loi introduite en Intelligence Artificielle, celle-ci suppose une gestion des exceptions. Par exemple :

Une tête a des yeux
les yeux servent à voir
mais * *la tête du chat voit la souris.*

En assimilant l'action du verbe au verbe, on peut aussi, pour éliminer la relation :

vente aux algues dans (2)

faire appel à la formule :

K(N) = K (V. p. coi)

utilisée par J.H. JAYEZ (70) et qui signifie que la catégorie sémantique du substantif N est une catégorie de complément d'objet indirect introduit par la préposition p possible pour le verbe V.

Dans le cas présent, la catégorie "plante aquatique" (*) ne satisfaisant pas l'égalité, la relation candidate est par conséquent rejetée.

Nous avons examiné jusqu'ici les difficultés que pose la reconnaissance des relations entre les substantifs. La reconnaissance des relations entre substantifs et adjectifs postérieurs pose les mêmes difficultés.

Considérons les chaines nominales suivantes :

(3) *Les réacteurs à neutrons rapides*
ARTD S PREPART S ADJP

(4) *Les vestes à doublures noires*
ARTD S PREPART S ADJP

(5) *La matrice de précédence fréquentielle*
ARTD S PREPART S ADJP

(6) *Le laboratoire d' automatique théorique*
ARTD S PREPART S ADJP

Là encore, les structures syntaxiques superficielles sont identiques (**) pourtant les représentations fléchées sont différentes.
On a :

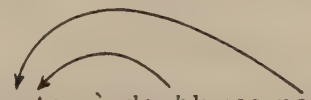
(*) Ce trait est pris dans le ROBERT.

(**) Comparer de même : { *Les soldats dédaigneux du danger.* / *Les révélations scandaleuses du journal.*

{ *Un élève attentif à la leçon.* / *Un thé chaud à la menthe.*

(3a) Les réacteurs à neutrons rapides
qui se flèche de la même manière que (6) ;

(6a) Le laboratoire d'automatique théorique



(4a) Les vestes à doublures noires
qui se flèche de la même manière que (5) ;

(5a) La matrice de précédence fréquentielle

Notons que dans tous ces cas, l'accord en genre et en nombre n'est d'aucun secours.

Pour résoudre ces problèmes d'ambiguïté de fléchage on peut faire appel à des règles similaires aux règles que nous avons vues ci-dessus, c'est à dire :

- soit des règles explicites autorisant les seules relations possibles ;
- soit des règles basées sur la compatibilité des traits sémantiques.

Cependant, il arrive que plusieurs interprétations soient possibles ; ainsi (4) pourrait accepter le flêchage suivant :

(4b) *Les vestes à doublures noires*

La résolution éventuelle de ces cas demande alors que l'on fasse intervenir un contexte plus large. Ces chaînes étant souvent plongées dans un corpus, nous verrons plus loin un moyen de tenir compte de ce contexte textuel (cf. ch. V, § IV.2).

II- Reconnaissance de la relation "Sujet - Verbe" :

La reconnaissance de la relation "Sujet - Verbe" en tant qu'agent pouvant accomplir l'action du verbe, présente des difficultés analogues à celles mentionnées plus haut, à savoir une ambiguïté due à la présence de plusieurs relations potentielles, conséquence de la non-contiguïté fréquente des éléments mis en relation.

Considérons les deux segments de phrases suivants :

(7) *La copie de l' arrêté autorisant ...*
ARTD S PREP ARTD S PPRES

(8) *Le ministre de la reconstruction agissant*
ARTD S PREP ARTD S PPRES

où PPRES désigne la catégorie grammaticale "verbe au participe présent".

Les deux segments ont même structure syntaxique mais se flèchent différemment :

(7a) *La copie de l'arrêté autorisant*

(8a) *Le ministre de la reconstruction agissant*

Là encore, l'usage de critères syntaxiques ne peut pas nous sortir d'affaire. On peut alors faire appel aux "restrictions" de M. SALKOFF (cité par JAYEZ) qui s'énoncent ainsi :

"Si une classe X a la valeur X_i, alors une autre classe Y, qui lui est grammaticalement liée doit avoir (ou ne pas avoir) la valeur Y_j" (SALKOFF, 105)

Ces restrictions, au nombre de 200 environ, permettent à M. SALKOFF de préciser l'acceptabilité des diverses combinaisons d'éléments à l'intérieur des chaînes.

Ainsi, dans une chaîne telle que :

N_1 *admirait le tableau*

seuls des noms tels que "homme" sont acceptables en N_1. M. SALKOFF a pu ainsi dégager 15 sous-classes de noms, 30 de verbes, 60 d'objets, 10 d'adjectifs et 20 d'adverbes (SALKOFF, 104).

Il subsiste évidemment toujours la solution implicite qui consiste à n'autoriser que les relations possibles ; dans notre exemple :

arrêté ⟶ *autoriser*
ministre ⟶ *agir*

lesquels auraient au préalable été stockées.

III- Reconnaissance de la relation "Verbe - Complément" :

Nous n'introduirons pas pour l'instant les différentiations traditionnelles faites au sujet du complément (JAYEZ, 70), notre propos étant toujours de montrer en quoi la reconnaissance de certaines relations demande une information sémantique.

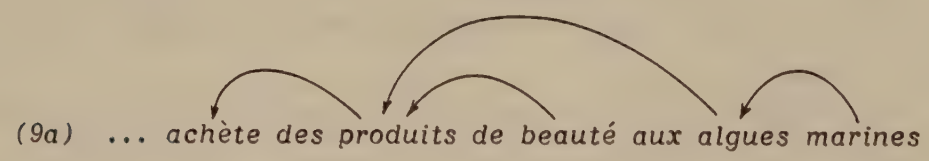
Considérons les deux segments de phrase suivants :

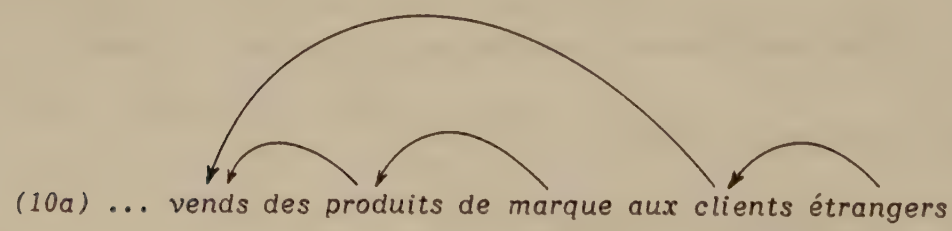
(9) .... *achète des produits de beauté aux algues marines*
INDC PREPART S PREPART S PREPART S ADJP

(10) .... *vends des produits de marque aux clients étrangers*
INDC PREPART S PREPART S PREPART S ADJP

où INDC désigne la catégorie grammaticale "verbe à l'indicatif".

Ces deux segments ont encore des structures syntaxiques identiques mais leur fléchage est différent :





La littérature peut nous offrir ici d'utiliser par exemple, les formules de WILKS. Celles-ci rappelons-le, sont basées sur des "unités sémantiques élémentaires". 60 unités ont été retenues pour être réparties en 5 classes (Y. WILKS, 112, p. 123) :

(a) entities: MAN(human being), STUFF(substances), THING(physical object), PART(parts of things), FOLK(human groups), ACT(acts), STATE(states of existence), BEAST(animals), etc.
(b) actions: FORCE(compels), CAUSE(causes to happen), FLOW(moving as liquids do), PICK(choosing), BE(exists) etc.
(c) type indicators: KIND(being a quality), HOW(being a type of action) etc.
(d) sorts: CONT(being a container), GOOD(being morally acceptable), THRU(being an aperture) etc.
(e) cases: TO(direction), SOUR(source), GOAL(goal or end), LOCA(location), SUBJ(actor or agent), OBJE(patient of action), IN(containment), POSS(possessed by) etc.

Ces unités sémantiques peuvent éventuellement être combinées (cela permet d'éviter d'allonger la liste des unités primitives).

Ex. :

(FLOW STUFF) exprime le caractère liquide

(THRU PART) indique une ouverture

En plus de ces 60 unités primitives, il existe des classes dont le nom commence par un astérisque.

Ex. :

**ANI : pour la classe des éléments animés MAN, BEAST, FOLK.*

**HUM : pour les éléments humains MAN et FOLK.*

La signification des mots est matérialisée par une formule utilisant ces divers éléments sémantiques. Par exemple, l'action "drink" sera exprimée par la formule :

```
((*ANI  SUBJ)(((FLOW  STUFF)OBJE)((*ANI  IN)(((THIS(*ANI
(THRU PART)))TO)(BE CAUSE)))))
```

Cette formule exprime une action effectuée en général par un sujet animé sur une substance liquide ; cette action amène le liquide à l'intérieur du sujet via une ouverture du sujet.

Une telle formule reste évidemment trop riche pour le propos qui nous intéresse ici.

IV- Reconnaissance des référents :

"1. pick up a big red block

OK

2. grasp the pyramid

I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN

3. find the block which is taller than the one you are holding and put it into the box

BY IT I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING

OK" (T. WINOGRAD, 113, p. 155)

On reconnait à travers ce dialogue l'univers des cubes et pyramides de WINOGRAD.

Cette expérience qui remonte à 1970 mettait déjà en oeuvre, pour résoudre les ambiguïtés inhérentes à la reconnaissance des référents des pronoms, concurremment trois composantes dont était pourvu le système, à savoir : une composante syntaxique, une composante sémantique et une capacité spécialisé de raisonnement.

Nous allons examiner quelques exemples en français.

Considérons les deux segments de phrases suivants :

(11) Le propriétaire d' une forêt qui vend ...
ARTD S PREP ARTI S ?SUJ INDC

(12) Le capitaine d' un navire qui coule ....
ARTD S PREP ARTI S ?SUJ INDC

où ARTI désigne la catégorie grammaticale "article indéfini" et ?SUJ le pronom relatif en position sujet.

Seule une composante sémantique nous permettra de reconnaître les antécédents des pronoms relatifs, qui sont dans des positions "géographiques" différentes par rapport aux pronoms, et par conséquent d'obtenir les représentations fléchées distinctes suivantes :

(11a) Le propriétaire d'une forêt qui vend

(12a) le capitaine d'un navire qui coule

Les exemples ici abondent tant il est vrai qu'il existe divers cas faisant appel, somme toute, à différents niveaux de connaissance.

Ainsi, dans :

(13) Le chien qui voit un sucre, monte sur le buffet.

(14) Il veut le manger

G.SABAH, après avoir éliminé certains référents candidats grâce à des règles d'exclusivité (101, p.105) fait appel aux traits sémantiques rattachés au thème me du verbe "manger"("solide" et généralement "comestible") pour lever l'ambiguïté sur le pronom préverbal "le" dont les antécédents potentiels sont "sucre"

et "buffet". Le premier trait sémantique ("solide") ne donne ici aucune indication mais le second permet d'éliminer "buffet". La seconde phrase est alors comprise comme :

(14') le chien veut manger le sucre.

S'il subsiste des ambiguïtés, G.SABAH fait alors intervenir des informations "pragmatiques" telles celles relatives aux habitudes alimentaires des chats, par exemple, pour les phrases :

(15) Dans le jardin, un chat voit une souris.

(16) Il se glisse derrière une laitue car il veut la manger

Si l'on considère le point comme délimiteur de phrase, il convient de noter, qu'ici plus qu'ailleurs, la reconnaissance de la référence des pronoms fait appel à une analyse qui sort souvent du cadre de la phrase.

Exemple :

"Dans les habitations des deuxième, troisième et quatrième familles, les gaines respectivement destinées à recevoir les colonnes montantes de gaz et d'électricité doivent être sans communication entre elles. Celles destinées aux colonnes montantes de gaz doivent être ouvertes à leur partie inférieure et communiquer avec l'atmosphère à leur partie supérieure.
Celles destinées aux colonnes montantes d'électricité doivent être fermées à leur partie supérieure."

(extrait de la règlementation de l'urbanisme)

Nous ne tenterons pas de montrer, aux travers d'études fort connues (SCHANK, WILKS, BOBROW, CHARNIAK, etc.) la nécessité des informations dites pragmatiques, quoique là encore, distinction et définition soient quelquefois difficiles à établir.

Citons toutefois ces deux exemples empruntés à J. PITRAT (95) qui font apparaître "la masse colossale d'informations pragmatiques".

(17) Essuyez bien les poissons afin d'enlever les écailles, coupez la tête, huilez-les très légèrement et mettez-les à cuire très rapidement sur un grill bien chaud.

(18) Le professeur envoya le cancre chez le censeur parce qu'il voulait :

1°/ lancer des boulettes ;

2°/ avoir la paix ;

3°/ le voir.

V. Reconnaissance de la polysémie :

Il est une autre forme d'ambiguïté qui peut être rapprochée des cas que nous venons d'évoquer, au sens où elle met en jeu pour sa résolution le même type d'informations.

Certains cas de polysémie (cf. ch. IV, § IV.3.3.1) peuvent être en effet considérés comme ayant plusieurs représentations (autant que de sens différents) en machine.

Par exemple, au mot VOL correspondra deux chaînes de caractères : VOL (OISEAU) et VOL (DEROBER) qui rappellent les deux principales significations attachées au mot VOL.

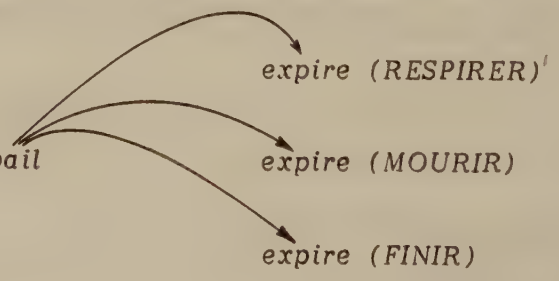
Notons que les expliciteurs (OISEAU et DEROBER dans l'exemple) peuvent être des marques sémantiques.

La représentation proposée ici est évidemment relativement fruste comparée aux formules de WILKS par exemple (cf. § III). Elle est cependant suffisante pour illustrer notre propos et montrer, encore une fois s'il était nécessaire, le caractère indispensable des relations sémantiques soit entre unités lexicales, soit entre unités lexicales et marques sémantiques.

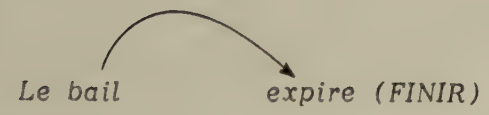
Considérons le segment :

(19) Le bail expire à la fin du mois

Si l'on retient pour "expirer" les trois acceptions suivantes, que l'on peut trouver dans un dictionnaire courant de la langue : RESPIRER, MOURIR et FINIR, (19) peut alors être représentée comme suit :



Il est clair que nous sommes en présence d'une ambiguïté qui peut être levée soit par la relation explicite :



soit par la compatibilité (ou non) des marques sémantiques associées à chacun des mots.

Notons que c'est ainsi que WILKS opère pour discriminer le sens du mot "crook" dans :

(20) The big policeman interrogated the crook.

Au cours de l'analyse, "crook" est rencontré avec deux sens possibles
- "escroc" spécifié par l'élément sémantique primitif MAN ;
- "houlette" par THING

La confrontation de ces éléments primitifs avec ceux attachés à l'objet du verbe "interrogated", (MAN OBJE) en l'occurrence, permettent au système de **préférer** le sens de "escroc".

Rappelons brièvement le système CSAW (Choosing the Senses of Ambiguous Words) de P.J. HAYES (cité par J.H. JAYEZ, 70) qui à partir d'une description des mots sous forme de sous-ensembles d'un réseau sémantique, tente de trouver le sens des mots dans le contexte où ils sont plongés.

Ainsi, dans :

(21) Fred walked into the room ; his arm was covered in bandages.

"arm" a plusieurs sens possibles (bras, arme).

Pour résoudre cette ambiguïté, P.J. HAYES (67) fait appel à une description du concept D - HUMAN qui est actualisé pour représenter Fred. Au cours de cette actualisation, arm est relié au noeud descripteur N - ARM et, attribuant une priorité à la description dernièrement actualisée, CSAW retient le sens de bras humain.

VI. Conclusion :

Outre la nécessité d'une composante sémantique, nous avons tenté de dégager, au travers des exemples que nous avons donnés, deux difficultés de nature différente, inhérentes aux solutions traditionnellement retenues :

- l'une, fondamentale, concerne la définition des traits ou marques sémantiques (*);
- l'autre, importante sur le plan technique, est propre à la construction matérielle du lexique unités morphologiques - traits sémantiques associés.

La corrélation de ces deux difficultés explique, nous semble-t-il, en partie le caractère généralement restreint des univers sur lesquels ont porté les expériences.

C'est la conjugaison de ces divers aspects qui nous a amené à envisager l'élaboration d'analyseurs syntaxico-sémantiques se fondant sur des méthodes d'acquisition automatique.

Afin de situer nos travaux par rapport à l'ensemble des recherches actuelles, nous remarquerons à nouveau que, pour la plupart, celles-ci se sont plus préoccupées de la définition des informations nécessaires et de leur représentation et utilisation que de leur acquisition.

C'est ainsi par exemple, qu'ont été introduites des notions, certes fondadamentales pour l'étude de la compréhension, telles que celles de "frames" (MINSKY ; BOBROW et WINOGRAD dans KRL) et qui découlent et renseignent sur la façon dont les évènements se déroulent généralement.

Les recherches qui ont développé des notions analogues : ABELSON ("plans") CHARNIAK ("démons"), COULON & KAYSER ("veilleurs"), SCHANK ("scripts"), WILKS (règles de sens commun) ont généralement considéré comme réalisées les étapes d'analyse syntaxico-sémantique, soit par un principe de "division du travail" (CHARNIAC, SCHANK) soit par des simplifications (WILKS).

Ceci est par exemple très marqué chez CHARNIAK, qui s'intéressant au problème de l'inférence, a délaissé volontairement toute phase d'analyse syntaxique en supposant que les chaînes du discours étaient a priori exprimées sous une forme sémantique interne.

Dès lors, nos recherches peuvent être considérées comme situées en amont par rapport à ces différentes tendances, dans la mesure où, n'émettant aucune restriction sur la nature des textes, nous nous proposons d'effectuer précisément une analyse syntaxico-sémantique sans apport préalable d'une information quelconque sur les compatibilités sémantiques.

(*) Nous considérons comme sous-entendu le caractère "flou" qui entoure les frontières entre syntaxe, sémantique et pragmatique et qu'il est par conséquent souvent difficile, dans certains cas, de savoir dans quel niveau descriptif on est. Aussi, et bien que l'on puisse aboutir à des classes distributionnelles, nous préférerons envisager celles-ci du point qui aurait pu permettre de les établir, à savoir sémantique.

Chapitre deuxième

NOTIONS ET HYPOTHESES DE BASE

Dans le chapitre précédent nous avons tenté de matérialiser, au travers des exemples que nous avons donnés, quelques difficultés sous jacentes à la reconnaissance automatique des différentes relations entre mots dans la phrase.

L'accent a porté cependant sur un seul aspect qui était celui de la nécessité d'une information sémantique.

Nous allons dans ce chapitre cerner un peu plus le problème qui nous préoccupe :

- en précisant davantage la nature et la forme que pourra revêtir cette information sémantique,
- en avançant les hypothèses qui sont à la base de notre système.

I. Notion de relation lexicale sémantique•(RLS) :

Nous allons tenter de donner une définition de ce terme, mais ce sont surtout les exemples qui illustreront la signification que nous attachons à cette notion.

Soit m_1 m_2... m_n une chaîne de mots pleins•extraits d'un texte T. On appellera RLS n-aire sur m_1 m_2 ... m_n une relation définie par la co-occurrence de ces mots dans une phrase et précisée par des observations syntaxiques qui expriment une réalité sémantique (*)

Exemples :

Soit :

T_1 = Ce maigre repas ...

on peut en extraire la RLS notée :

(maigre, repas, ADJ-ANT)

qui est différente de la RLS que l'on peut extraire de :

T_2 = Un repas maigre ...

et que nous noterons :

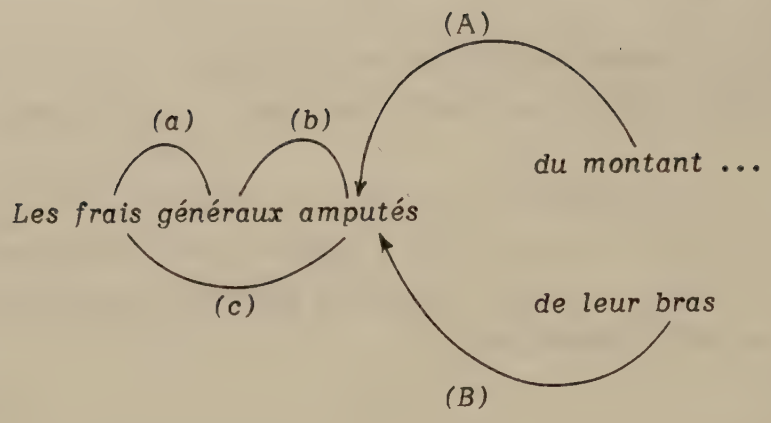
(maigre, repas, ADJ-POST).

On voit donc que des phénomènes syntaxiques peuvent traduire des faits sémantiques (**)

En général, il semble nécessaire de connaître diverses relations d'ordre quelconque pour résoudre les polysémies et ambiguïtés structurales. Mais nous avons vu au chapitre précédent, que la plupart des situations rencontrées ne demandent que la connaissance de relations binaires. Il nous paraît même difficile de trouver des exemples faisant appel à un ordre supérieur à deux.
Voici néanmoins un exemple où l'ordre est trois :

(*) "... le structural exprime le sémantique", L. TESNIERE, 108, p.42

(**) "Nous pourrions ici justifier le choix de ce mode de définition, en soulignant qu'il est parfaitement compatible avec une vision d'ensemble de la linguistique selon laquelle la pragmatique produit la sémantique, qui à son tour détermine la syntaxe ; en effet, dire que A produit B n'interdit pas de faire découler la connaissance de A de celle de B. Ceci est expliqué en détail dans Conan Doyle, qui fait dire à Sherlock Holmes que, bien évidemment, c'est le criminel qui produit l'empreinte, et non l'inverse, mais la connaissance de l'empreinte peut conduire à la connaissance du criminel". (J.B. BERTHELIN, 19) .



pour détruire (b) ou (c) en utilisant (A) ou (B) il faut une relation ternaire (pour d'autres exemples, voir conclusion générale, § I).

Ceci nous amène donc à formuler notre première hypothèse :

La connaissance de relations lexicales sémantiques d'ordre 2 peut suffire pour résoudre , dans une grande proportion, les ambiguïtés inhérentes à la reconnaissance des différentes relations de dépendance qui sous-tendent le discours.

Cette hypothèse appelle cependant deux remarques :

a) **Validité** : Les RLS d'ordre 2 (ou plus) figurant dans les corpus textuels traités seront considérées comme valides. On entend ainsi que l'on ne présentera pas à l'analyse des phrases comme par exemple :

Le silence vertébral indispose la voile licite
(L. TESNIERE, 108, p. 41)

sous peine de voir répertorier comme valides des relations illicites telles que *"le silence vertébral"* ou *"la voile licite"* .

b) **Complétude** : Si la prise en compte exclusive des RLS binaires semble a priori suffisante et en tout cas plus aisée, elle ne s'opère cependant pas sans que l'on ait à s'accommoder d'une certaine distorsion.

En effet, si l'on considère les RLS binaires indépendamment du contexte d'où elles ont été extraites, on peut aboutir à des relations fausses, et ce tant du point de vue génératif que du point de vue de l'analyse.

Ainsi, s'il est vrai que l'on peut *"jouer avec les mots"* et qu'un *"chat peut jouer"* , il n'est pas vrai qu'un *"chat joue avec les mots"* ou *"au scrabble"*.

Inversement, s'il est vrai qu'une *"maison tremble en cas de séïsme"* , on ne peut en inférer que la relation binaire *"la maison tremble"* soit une relation acceptable pour *"maison"* dans le cas général.

Cependant, étant donné la manière dont notre système fait usage de ses connaissances, c'est-à-dire : en fonction des besoins qui apparaissent dans le contexte ; une connaissance manifestement fausse a toutes les chances de n'être jamais sollicitée, ce qui "neutralise" l'erreur qui conduit à son acquisition.

Cette hypothèse constitue l'une des hypothèse de base du système que nous présentons.

Nous appellerons donc dans la suite de l'exposé RLS simplement le triplet (m_1, m_2, R), qu'il nous arrivera de noter également (m_1, R, m_2).

Rappelons que nous avons laissé apparaître au travers du chapitre précédent notre choix d'orientation vers une composante sémantique explicite, c'est à dire basée sur des relations de compatibilité sémantique entre mots du lexique. Les relations de dépendance vues alors constituent des relations lexicales sémantiques.

Ainsi, si nous désignons par :

- *SUJ* la relation de "sujet",
- *COI* la relation de "complément d'objet indirect",
- *Q* la relation de "qualification"(*) ;

l'ensemble des triplets suivants :

(a) (*vendre, COI, client*)
(b) (*neutrons, Q, rapides*)
(c) (*agir, SUJ, ministre*)

constituent des relations lexicales sémantiques.

Ce sont là des informations types qui peuvent nous permettre de résoudre les ambiguïtés syntaxiques dont nous avons vu quelques spécimens dans le chapitre précédent. Mais rappelons encore qu'il ne s'agit pas de les rassembler par des moyens "manuels", mais au contraire de les obtenir automatiquement à partir des textes.

Si l'on reprend les exemples (1), (3) et (8) du chapitre I, auxquels nous pouvons appliquer respectivement, pour leur résolution, les RLS (a), (b) et (c), on remarque la nécessité d'un autre type d'information dont nous n'avons pas encore parlé.

(*) Il faut remarquer que ces catégories sont plus larges que celles que recouvre un "cas sémantique" habituel.

En effet, outre les relations de "quasi-identité" que l'on voit exister entre *"clients"* et *"client"* ou *"agissant"* et *"agir"*, on remarque que si l'on veut appliquer avec succès la RLS (a) à l'exemple (1), on doit pouvoir établir une relation particulière entre *"vente"* et *"vendre"* (nous ne ferons pas allusion aux transformations syntaxiques nécessaires à ce stade de l'exposé).

L'aspect général de la définition proposée pour la notion de RLS permet de répertorier comme telle ce deuxième type d'information. On y trouvera donc des relations de synonymie, d'antonymie, de nom d'agent, etc. (A.K. ZHOLKOVSKIJ, I.A. MEL'CHUK, 115)

Si l'on désigne par :

- *SYN* la relation de "synonymie";
- *ANT* la relation d'"antonymie";
- *AG* la relation "nom d'agent";
- *ACT* la relation "nom d'action";

l'ensemble des triplets suivants :

(logiciel, SYN, software) ; (acheter, AG, acheteur)
(lenteur, ANT, rapidité) ; (afficher, ACT, affichage)

constitueront également des RLS.

Cette définition générale n'est pas sans rappeler *"les relations lexico-sémantiques ... qu'il est nécessaire de connaître pour obtenir des périphrases synonymes et pour donner à la traduction son caractère idiomatique"* de A.K. ZHOLKOLVSKIJ et I.A. MEL'CHUK (115) à qui nous emprunterons également les notions de *"fonction lexicale"* (nous dirons plutôt correspondance lexicale) et de *"corrélatif lexical"* . On associe à la *"relation entre* (m_1) *et son corrélatif* (m_2) *une fonction lexicale dans laquelle* (m_1) *est l'argument et le corrélatif* (m_2) *la valeur"*.

On peut alors écrire :

COI (vendre) = client
Q (neutron) = rapide
SUJ (agir) = ministre

SYN (logiciel) = software
ANT (lenteur) = rapidité ;
AG (acheter) = acheteur;
ACT (afficher) = affichage .

La construction en partie automatique de ce deuxième type d'information sera traitée plus loin, dans le chapitre IV.

II. RLS paradigmatiques - RLS syntagmatiques :

Pour distinguer les deux types d'information dont on vient de voir quelques exemples, nous utiliserons la terminologie traditionnelle en appliquant le qualificatif syntagmatiques aux premières RLS, et le qualificatif paradigmatiques aux dernières RLS.

Nous aurons donc :

RLS syntagmatiques	RLS paradigmatiques
COI(vendre) = client	*SYN(logiciel) = software*
Q (neutron) = rapide	*ANT(lenteur) = rapidité*
SUJ(agir) = ministre	*AG(acheter) = acheteur* (*)
	ACT(afficher) = affichage (*)

Nous distinguerons donc deux classes de correspondances lexicales : paradigmatiques et syntagmatiques, dont les valeurs seront respectivement des corrélatifs lexicaux paradigmatiques et syntagmatiques.

Notons que ceux-ci peuvent être confondus. Ainsi :

AG(acheter) = acheteur
ou encore *= client*

(*) Il serait gênant, pour un système de synthèse, de dire "l'acheteur du juge" ou "l'affichage de sentiments". Mais en analyse, le fait de ne jamais rencontrer ces constructions supprime le problème.

Or "client" est un corrélatif lexical syntagmatique de "acheter" (cf. ch.III, fig. 46(b)).

Cependant, lorsqu'il n'y aura pas d'ambiguïté, nous omettrons ces qualificatifs afin de ne pas alourdir l'exposé.

Remarquons dès maintenant que la dichotomie ainsi pratiquée dans l'ensemble des RLS laisse clairement se dégager deux axes dans la représentation des connaissances. On peut d'ailleurs noter qu'en 1967, A.K. ZHOLKOVSKIJ et I.A. MEL'CHUK avaient déjà montré, dans le cadre de la synthèse sémantique (système de périphrase) (115) , la nécessité de telles informations et préconisé une telle organisation de ces informations dans un dictionnaire : *"Ils (les articles du dictionnaire) contiennent aussi, et ceci est une innovation, toutes les fonctions lexicales du mot : en premier lieu, ses variantes "paradigmatiques", c'est à dire les moyens qui permettent de remplacer* i_o *(mot clé) dans des contextes définis et sous des conditions déterminées ; en second lieu, dans la mesure du possible, tous les renseignements sur l'expression idiomatique des sens qui lui sont syntagmatiquement liés"* (115, p.29) .

III. RLS syntagmatiques :

III. 1. Introduction :

Nous allons dans ce qui suit énumérer les principales RLS syntagmatiques que nous tenterons d'obtenir.

Rappelons que celles-ci ont pour but d'apporter une composante qui permettra la résolution des ambiguïtés structurales. Leurs définitions seront donc issues de l'analyse du problème telle que nous l'avons esquissée dans le premier chapitre. Elles sont par conséquent opératoires et leur justification se fera a posteriori, c'est à dire au vu des résultats obtenus.

On peut cependant se faire une idée de ce qu'elles peuvent être, en représentant, dans un schéma (cf. fig. 1.) certaines transitions (*) possibles entre les quatre catégories de mots pleins.

(*) Nous donnerons un nom à chacune de ces transitions au fur et à mesure que l'on avancera dans l'exposé.

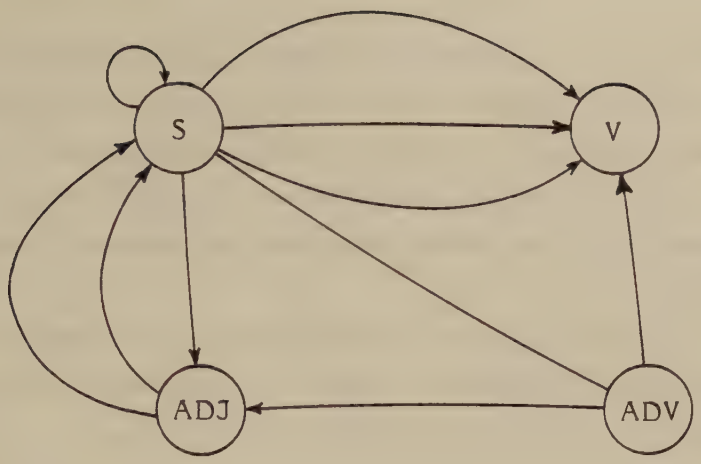


fig. 1.

Remarques :

1. Les transitions sont, rappelons-le, orientées : c'est à dire que l'on fait une distinction entre par exemple :

(1) *Pierre mange une poule*

et

(2) *Une poule mange*

2. Il peut exister plusieurs transitions, dans le même sens, entre deux sommets exemple :

(3) *Pierre mange avec une amie*

à ne pas confondre avec(1).

3. Il résulte de la précédente remarque que le nombre de transitions, et donc de RLS, n'est pas limité par le nombre de sommets.

Pour la clarté de l'exposé, mais aussi parce que nous la retrouverons loin, au niveau algorithmique (cf. ch.III, §II. 2), nous introduirons une distinction entre relations internes à une même chaîne (**relations homosyntagmatiques**) et relations s'établissant entre éléments appartenant à des chaînes diférentes (**relations hétérosyntagmatiques**).

Cette distinction, bien qu'ayant une réalité algorithmique, ne sera cependant que "virtuelle", le but étant de mettre l'accent sur l'indépendance entre RLS syntagmatiques d'une part et structures de surface qui leur donnent naissance d'autre part. C'est pourquoi, nous serons amenés, chaque fois qu'il sera possible, à rassembler sous une même étiquette des relations hétéro et homosyntagmatiques.

Nous verrons en effet que cette indépendance entre les RLS et les structures superficielles est nécessaire si l'on veut accroître les capacités inférentielles du système (cf. ch.V).

III. 2. Description des RLS syntagmatiques :

Une correspondance lexicale est une relation binaire. Chaque RLS syntagmatique sera donc décrite selon le modèle suivant :

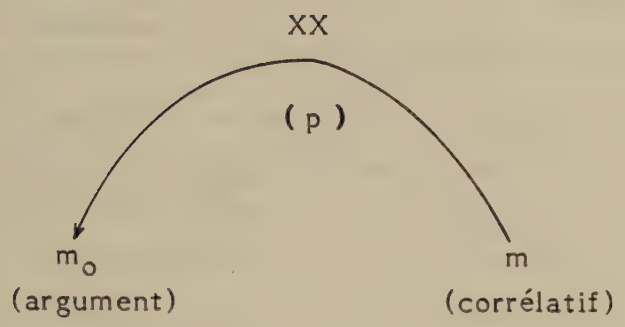
a) Définition : Dans cette rubrique nous spécifierons :
- Le nom de la correspondance lexicale associée. Celui-ci sera constitué de deux lettres représentant l'abréviation de la RLS, éventuellement indicée par le lien (copule ou préposition) reliant les deux termes.
- Les ensembles de départ et d'arrivée notés respectivement E_D (pour l'argument de la correspondance) et E_A (pour le corrélatif lexical). Ils correspondent à des sous-ensembles du lexique dont les éléments sont les mots possédant une (ou plusieurs) catégorie(s) grammaticale(s) donnée(s).
- Eventuellement la nature du lien.
- La nature de la relation.

b) Exemples.

c) Commentaires.

Remarque:

Sur le plan graphique, nous utiliserons le plus souvent une représentation fléchée. La liaison sera, par convention, matérialisée au moyen d'une flèche orientée du corrélatif vers l'argument. Celle-ci sera étiquetée par le nom de la correspondance lexicale qu'elle représente sous la forme suivante :



XX : nom générique de la RLS

p : nom spécifique du lien ou classe de liens.

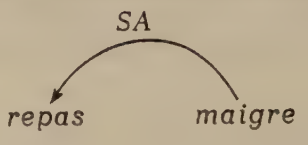
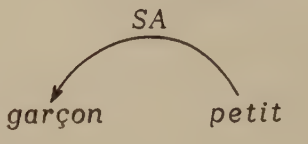
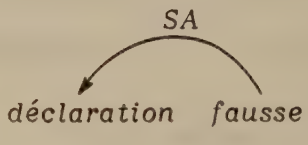
fig. 2.

III. 2. 1. RLS Substantif - adjectif Antérieur :

a) **Définition :**

- Nom symbolique de la correspondance lexicale associée : SA
- E_D : Ce sont les mots du lexique admettant la catégorie grammaticale S (Substantif). Nous noterons **S** cet ensemble.
- E_A : Ce sont les mots du lexique admettant la catégorie grammaticale ADJA (ADJectif Antérieur). Ils constituent un sous-ensemble de l'ensemble des adjectifs du lexique.
- La correspondance lexicale SA donne la relation sémantique qui s'établit entre l'adjectif qualificatif et le substantif auquel il se rapporte dans une structure dans laquelle l'adjectif précède le substantif.

b) **Exemples :**

N°	Structures de surface	RLS
1	... *un maigre repas* ...	
2	... *un petit garçon* ...	
3	*En cas de fausse déclaration* ...	

c) **Commentaires :**

- C'est à dessein que nous ne retrouvons pas sous la même dénomination les RLS Substantif - Adjectif antérieur et Substantif - Adjectif postérieur (cf. § III.2.2). En effet, si souvent un adjectif donné peut se placer avant ou après un substantif sans que soit sensiblement affecté le sens (*une jolie robe, une robe jolie*), souvent il arrive également que le sens soit affecté : *maigre repas, repas maigre* . Il est enfin des adjectifs qui ne peuvent se placer que avant ou après un substantif donné :

un tube métallique — * *un métallique tube*

Le code civil — * *Le civil code*

Le vingtième siècle — * *Le siècle vingtième*

La possibilité d'établir une équivalence entre ces deux types de relations (Substantif - Adjectif antérieur et Substantif - Adjectif postérieur) ne sera cependant pas écartée (cf. ch. II, § IV, ch. V, § III et IV).

- La correspondance lexicale SA est homosyntagmatique : les RLS qu'elle engendre ne proviennent en effet que de chaînes nominales.

III. 2. 2. RLS Substantif - adjectif Postérieur :

a) **Définition :**

- Correspondance lexicale associée : SP
- E_D : C'est l'ensemble **S** des substantifs du lexique.
- E_A : C'est l'ensemble A_S des adjectifs du lexique.
- La correspondance lexicale SP donne la relation sémantique qui s'établit entre l'adjectif et le substantif auquel il se rapporte dans une structure où le substantif précède l'adjectif ou encore, dans une structure où l'adjectif est attribut.

b) **Exemples :**

N°	Structures de surface	RLS
1	*Le code civil*	SP : *codel* — *civil*
2	*Le tube métallique ...* *Le tube est métallique ...*	SP : *tube* — *métallique*
3	*Le transport bénévole ...*	SP : *transport* — *bénévole*
4	*Les exemptions applicables ...* *Les exemptions sont applicables ...*	SP : *exemption* — *applicable*

c) **Commentaires :**

- Les exemples choisis montrent que la correspondance lexicale SP définit un ensemble de RLS non-homogène. En effet, si la RLS (1) (qui constitue la RLS type) ne peut revêtir qu'une seule forme en surface les relations (2), (3) et (4) peuvent revêtir d'autres formes en surface. Ainsi, on peut trouver respectivement :

Le tube de métal

ou encore *Le tube en métal* *pour (2)*

transporter bénévolement *pour (3)*

et enfin *appliquer les exemptions* *pour (4)*

Nous serons donc amenés, si l'on veut accroître la composante inférentielle du système, à établir certaines équivalences entre des RLS particulières (cf. ch. II, § IV).

Dans le cas présent, les "synonymies" suivantes devront pouvoir être établies:

Certaines RLS SP avec certaines RLS SS_p pour (2) (cf. § III. 2. 3) ;

d'autres RLS SP avec certaines RLS VA pour (3) (cf. § III. 2. 7) ;

enfin, certaines RLS SP avec certaines RLS VD pour (4) (cf. § III. 2. 5).

Un tableau des principales équivalences qui peuvent être établies entre correspondances lexicales est donné figure 3 .

Notons que nous ne disposons pas de règles transformationnelles et que ces "synonymies" ne peuvent être en tout état de cause que potentielles et non absolues. Nous verrons d'ailleurs qu'elles seront établies dynamiquement (cf. ch. V, § III et IV).

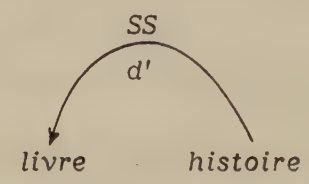
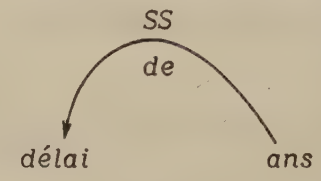
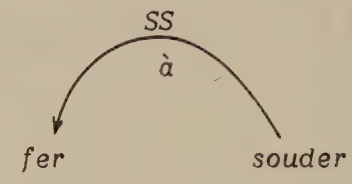
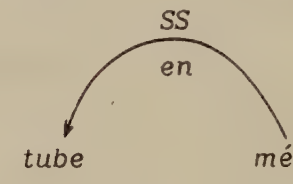
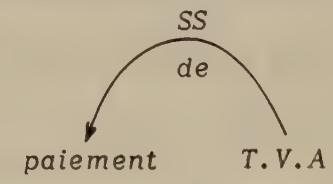
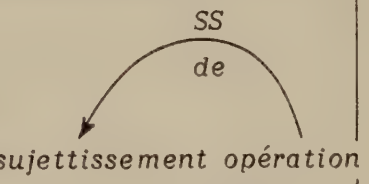
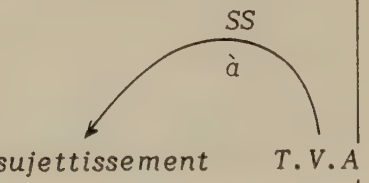
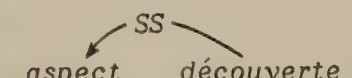
- Les RLS SP ainsi définies restent donc, malgré les exemples (2) et (3), essentiellement homosyntagmatiques.

III. 2. 3 RLS Substantif - Substantif :

a) **Définition :**

- Correspondance lexicale associée : SS_p
- E_D : Ensemble des substantifs du lexique.
- E_A : Ensemble des substantifs et infinitifs du lexique.
- Nature du lien : La RLS est ici souvent caractérisée par la présence d'une préposition, l'indice p représente la préposition qui supporte la relation Il prend ses valeurs dans l'ensemble **P** des prépositions, qui inclut certaines locutions prépositives. Il peut être absent (construction directe, exemples : *Le Président X, L'aspect découverte)*
- La correspondance lexicale SS_p donne la relation sémantique qui s'établit entre le substantif et l'un de ses substantifs ou infinitifs compléments par l'intermédiaire de la préposition p, ou par construction directe.

b) **Exemples :**

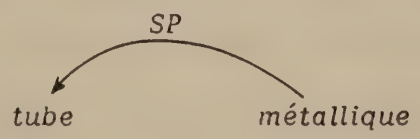
N°	Structures de surface	RLS
1	*Le livre d'histoire*	
2	*Dans un délai de trois ans*	
3	*Le fer à souder est sur la table.*	
4	*Le tube en métal* *Le tube est en métal*	
5	*L'acceptation du ministre*	
6	*Le paiement de la T.V.A.*	
7	*L'assujettissement de ces opérations à la T.V.A.*	
8		
9	*L'aspect découverte*	

c) **Commentaires :**

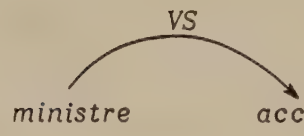
Les exemples choisis appellent des remarques analogues à celles qui se rapportent aux RLS précédentes.

- Nous remarquons tout d'abord que les RLS choisies ne sont pas toutes de même nature. De plus, elles peuvent en général revêtir différentes formes en surface.

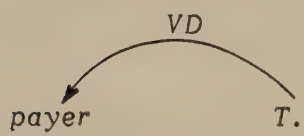
  Ainsi, comme nous l'avons vu plus haut (cf. § III. 2. 2), la RLS (4) pourrait revêtir la forme :



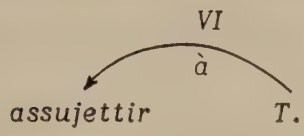
De même, les RLS (5), (6) et (8) peuvent être respectivement identifiées aux relations :



(cf. § III. 2. 4)



(cf. § III. 2. 5)



(cf. § III. 2. 6)

Outre la "transformation" potentielle :

$$SS_p \dashrightarrow SP$$

il y a lieu d'établir également les transformations potentielles suivantes :

$SS_p \dashrightarrow VS$ (cf. § III. 2. 4)
$SS_p \dashrightarrow VD$ (cf. § III. 2. 5)
$SS_p \dashrightarrow VI_p$ (cf. § III. 2. 6)

- Notons que celles-ci, même établies a priori, ne peuvent à elles seules suffire pour éviter cette hétérogénéïté apparente qui existe dans l'ensemble des RLS SS_p.

Les segments suivants :

réparation du garagiste
réparation du véhicule
ou même *réparation du robot*

identiques quant à leur structure syntaxique, montrent que dans le premier cas on est en présence d'une synonymie RLS SS_p ~ RLS VS alors que dans le deuxième cas on est en présence d'une synonymie RLS SS_p ~ RLS VD, le troisième étant ambigu.

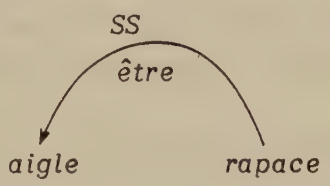
Le recours à des caractéristiques sémantiques attribuées aux mots mis en relation peut là encore nous sortir d'affaire. Mais, rappelons-le, nous ne disposons pas a priori de telles informations. Nous verrons que nous pourrons suppléer à ce manque en faisant appel à ces possibilités de transformations d'une part et aux RLS syntagmatiques et paradigmatiques que nous pourrons acquérir d'autre part (cf. ch. V, § III et IV).

Nous nous accommoderons donc pour l'instant de cette disparité dans l'ensemble des RLS définies par la correspondance lexicale SS_p.

- Les relations qui s'établissent entre substantifs dans des segments tels que :

l'aigle est un rapace

d'où l'on peut tirer la RLS :



ne feront pas partie de l'ensemble des relations LS définies par la correspondance lexicale SS_p.

Elles correspondent en effet à des relations dites de "définition" ou "générique - Spécifique". Nous les considérerons donc comme étant des relations paradigmatiques (cf. ch. IV). Notons que nous retrouvons là le problème des "frontières" entre relations syntagmatiques et relations paradigmatiques ou plutôt de la similitude de ces RLS quant à la forme qu'elles peuvent revêtir.

- Les RLS SS_p sont donc généralement homosyntagmatiques.

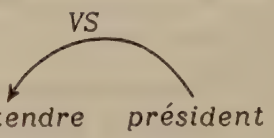
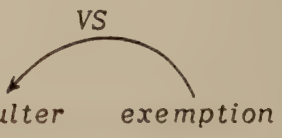
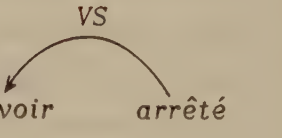
III. 2. 4. RLS Verbe - Sujet :

a) **Définition :**

- Correspondance lexicale associée : VS
- E_D : Ensemble **V** des mots du lexique admettant la catégorie grammaticale V (Verbe).
- E_A : Ensemble des substantifs du lexique.
- La correspondance lexicale VS donne la relation sémantique qui s'établit entre

le verbe et le sujet de la phrase de structure profonde (sujet à la forme active, complément d'agent à la forme passive).
Dans notre terminologie, *sujet* correspondra à l' *agent* ou au *siège du procès* (WAGNER & PINCHON, 111, p. 219) .

b) **Exemples :**

N°	Structures de surface	RLS
1	*Le président de la conférence peut entendre...*	
2	*Les exemptions de permis de construire résultant*	
3	*La décision est prise par le ministre*	
4	*Les exemptions prévues par le présent arrêté*	

c) **Commentaires :**

- Les exemples choisis montrent que les RLS VS peuvent revêtir diverses formes en surface.
- Rappelons que certaines relations SS_p devront être rapportées à l'ensemble des relations VS (cf. § III. 2. 3.).
- Il convient de souligner que la notion de sujet qui est ici mise en oeuvre correspond à la notion de sujet profond de la grammaire générative.

III. 2. 5. RLS Verbe - complément Direct :

a) **Définition :**

- Correspondance lexicale associée : VD
- E_D : **V**
- E_A : **S**
- La correspondance lexicale VD met en relation soit le verbe à la forme active et son complément d'objet direct, soit le verbe à la forme passive et son sujet, soit enfin le participe passé adjectif et le substantif auquel il se rapporte.

b) **Exemples :**

N°	Structures de surface	RLS
1	*Les exemptions prévues par le présent arrêté ne sont pas applicables aux travaux concernant les constructions ...*	VD : *prévoir* — *exemption*
2		VD : *concerner* — *construction*
3	*... ces ouvertures ne soient pas facilement obturées*	VD : *obturer* — *ouvertures*
4	*... toute personne peut consulter les pièces*	VD : *consulter* — *pièces*

c) **Commentaires :**

- Ici aussi, les RLS VD peuvent revêtir diverses formes en surface ainsi que le montrent les exemples.
- Certaines RLS SP ou SS_p devront être rapportées à l'ensemble des RLS définies par la correspondance lexicale VD (cf. § III. 2. 3. et III. 2. 4.).

III. 2. 6. RLS Verbe - complément Indirect :

a) **Définition :**

- Correspondance lexicle associée : VI_p
- E_D : **V**
- E_A : **S**
- Nature du lien : la RLS est ici caractérisée par la présence d'une préposition p. Indice p : voir § III. 2.3.
- La correspondance lexicale VI_p donne la relation qui peut s'établir entre le verbe (m_o) et le substantif (m) complément d'objet indirect ou complément circonstanciel.

b) **Exemples :**

N°	Structures de surface	RLS
1	*... ne sont pas applicables aux travaux ...*	*VI* — *aux* — *appliquer* — *travaux*
2	*Les conduits adossés à un mur ...*	*VI* — *à* — *adosser* — *mur*
3	*... qui ne sont pas assujettis à la procédure*	*VI* — *à* — *assujettir* — *procédure*

c) **Commentaires :**

Les remarques précédentes peuvent être reprises.

- Il y a une relative indépendance des RLS VI_p vis à vis des structures de surface qui leur donnent naissance.

- Certaines RLS SS_p devront être rapportées à l'ensemble des relations définies par VI_p (cf. § III. 2. 3.)
- Les expressions *"complément direct" et "complément indirect"* sont ici employées volontairement.

  En effet ces expressions *"ont rapport non à la nature du complément, mais simplement à sa construction"* (GREVISSE, 62, p. 103) . Or dans un cas comme dans l'autre, nous ne saurions faire les distinctions qui s'imposent quant à la nature du complément :

  1°/ *"Un complément est direct quand il se rattache directement, c'est à dire sans mot-outil, au mot complété ; il peut exprimer l'objet, mais aussi tout autre chose que l'objet : ce vin sent LE BOUCHON ; il sort LE SOIR."* (GREVISSE, 62, p. 103) .

  Dans :

  (1) *La souris mange le fromage*

  (2) *La souris mange du fromage*

  "fromage" est complément direct dans (1) mais sera complément indirect dans (2) si "du" est analysé "préposition article" et non "article partitif". Ceci nous amènera donc à admettre l'équivalence potentielle :

$$VD \sim VI_p$$

  2°/ Les distinctions entre *"complément indirect" et "complément circonstanciel"* sont également difficiles à établir : *"Dans des phrases comme :*

Qui donne au pauvre prête à Dieu.

Je cède la place à mon successeur.

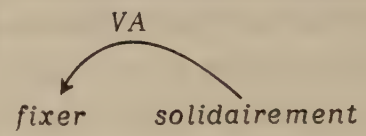
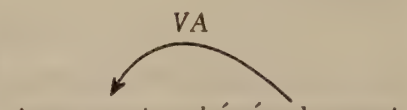
les mots "pauvre", "Dieu", "successeur" sont parfois rangés parmi les compléments circonstanciels et appelés compléments d'attribution" (GREVISSE, 62 p. 103) .
- Aussi il n'est pas introduit de sous-classification, au sens de la grammaire traditionnelle, dans l'ensemble des RLS VI_p. Chaque RLS est cependant particularisée au moyen de la préposition qui a permis de l'établir.

III. 2. 7. RLS Verbe - Adverbe :

a) **Définition :**

- Correspondance lexicale associée : VA
- E_D : C'est l'ensemble **V**
- E_A : C'est l'ensemble A_V des adverbes du lexique.
- La correspondance lexicale VA donne la relation sémantique qui s'établit entre le verbe et l'adverbe qui le modifie.

b) **Exemples :**

N°	Structures de surface	RLS
1	*... doivent être solidairement fixées ...*	
2	*... sont équipées de manière à fermer hermétiquement*	
3	*... transportées bénévolement ...* *... sont transportables bénévolement ...* *... bénévolement transportées ...* *... sont bénévolement transportables ...*	

c) **Commentaires :**

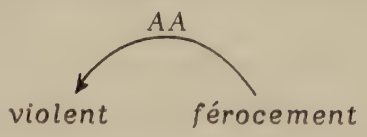
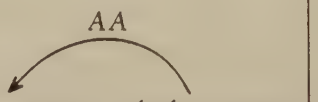
- Les RLS VA peuvent revêtir diverses formes en surface. Elles sont cependant toujours homosyntagmatiques : elles sont issues soit de chaînes verbales soit de chaînes nominales.
- Certaines relations AA (cf. § III. 2.8.) devront être rapportées à l'ensemble des relations définies par VA. D'une façon plus générale voir fig. 3 ou 4.

III. 2. 8. RLS Adjectif - Adverbe :

a) **Définition :**

- Correspondance lexicale associée : AA
- E_D : C'est l'ensemble A_S des adjectifs du lexique.
- E_A : C'est l'ensemble A_V des adverbes du lexique.
- La correspondance lexicale AA donne la relation sémantique qui s'établit entre l'adjectif (m_o) et l'adverbe (son corrélatif) qui le modifie.

b) **Exemples :**

N°	Sturctures de surface	RLS
1	... *férocement violent* ...	
2	*Les victimes transportables bénévolement* ...	
	Les victimes sont transportables bénévolement	

c) **Commentaires :**

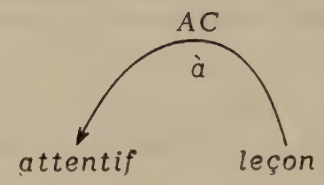
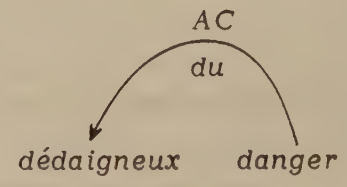
- Ces relations sont toujours internes à des syntagmes ou nominaux ou verbaux elles sont par conséquent homosyntagmatiques.

III. 2. 9. RLS Adjectif - Complément d'adjectif :

a) **Définition :**

- Correspondance lexicale associée : AC_p
- E_D : A_S
- E_A : S
- Nature du lien : là encore, la RLS est caractérisée par la présence d'une préposition. Indice p : voir § III. 2. 3.
- La correspondance lexicale AC_p donne la relation sémantique qui s'établit entre l'adjectif (m_o) et le substantif complément (m) introduit par la préposition p, ou par construction directe : *vert bouteille*.

b) **Exemples :**

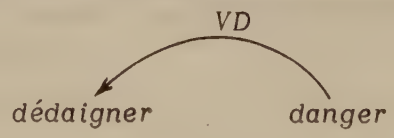
N°	Structure superficielle	RLS
1	*Les élèves attentifs à la leçon*	
2	*Les soldats dédaigneux du danger*	

c) **Commentaires :**

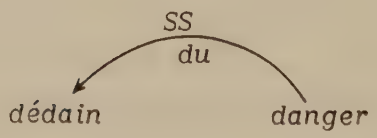
- La correspondance lexicale AC_p est homosyntagmatique : elle n'opère en effet que dans des chaînes nominales.
- Là aussi, il y a lieu d'établir certains liens synonymiques entre RLS provenant de structures superficielles différentes.

  Exemple :

  La RLS (2) pourrait être rapportée à :



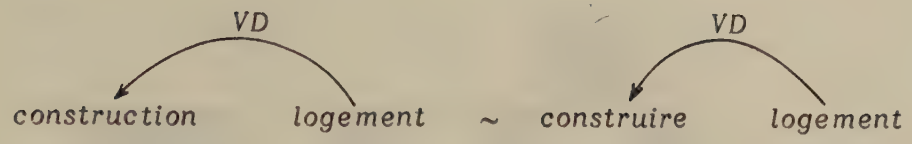
ou encore à :



IV. Synonymie lexico-syntaxique.

Certaines RLS équivalentes peuvent, comme nous l'avons vu, revêtir diverses formes et par conséquent être définies par différentes correspondances lexicales.

Exemple :



Nous dirons, en reprenant la terminologie usuelle, que nous sommes en présence d'une synonymie syntaxique ou plus généralement lexico-syntaxique.

La reconnaissance de cette synonymie est essentielle si l'on veut conférer au système une capacité inférentielle un tant soit peu importante.

La reconnaissance ou la construction volontaire (en régime de génération) de cette synonymie syntaxique se fait généralement par le recours à des règles de transformation définies à l'avance (cf. J.H. JAYEZ, 70 ; A.K. ZHOLKOVSKIJ I.A. MEL'CHUK, 115 ; B. VAUQUOIS, 109).

Notre approche sera, là encore, différente. Nous essaierons en effet de les simuler voire de les apprendre (cf. ch. V, § III et IV). Aussi, ne donnerons nous pour l'instant que l'ensemble des divers liens qui peuvent être établis entre les différentes correspondances lexicales que nous avons énumérées.

Une représentation matricielle de ces divers liens est donnée figure 3. Une case sombre (ou hachurée) signifie que l'on peut passer d'une RLS définie par une correspondance lexicale ligne (resp. colonne) à une RLS équivalente définie par une correspondance lexicale colonne (resp. ligne)

L'exemple donné ci-dessus nous montre que l'on peut passer d'une RLS définie par SS_p à une RLS équivalente définie par VD et inversement.

Par abus de langage, nous appellerons transformation syntaxique l'opération qui à une correspondance lexicale x fait correspondre une correspondance lexicale y.

La nature de ces transformations sera détaillée plus loin (cf. ch. V, § III & IV).

La matrice (cf. fig. 3) donnant l'ensemble des différentes synonymies syntaxiques possibles représente donc également l'ensemble potentiel des différentes transformations syntaxiques qui pourront être effectuées. Elle n'est cependant pas fournie au système. Nous verrons que les relations qu'elle renferme et que nous expliciterons plus loin (cf. ch. V, § III & IV) peuvent être construites par apprentissage.

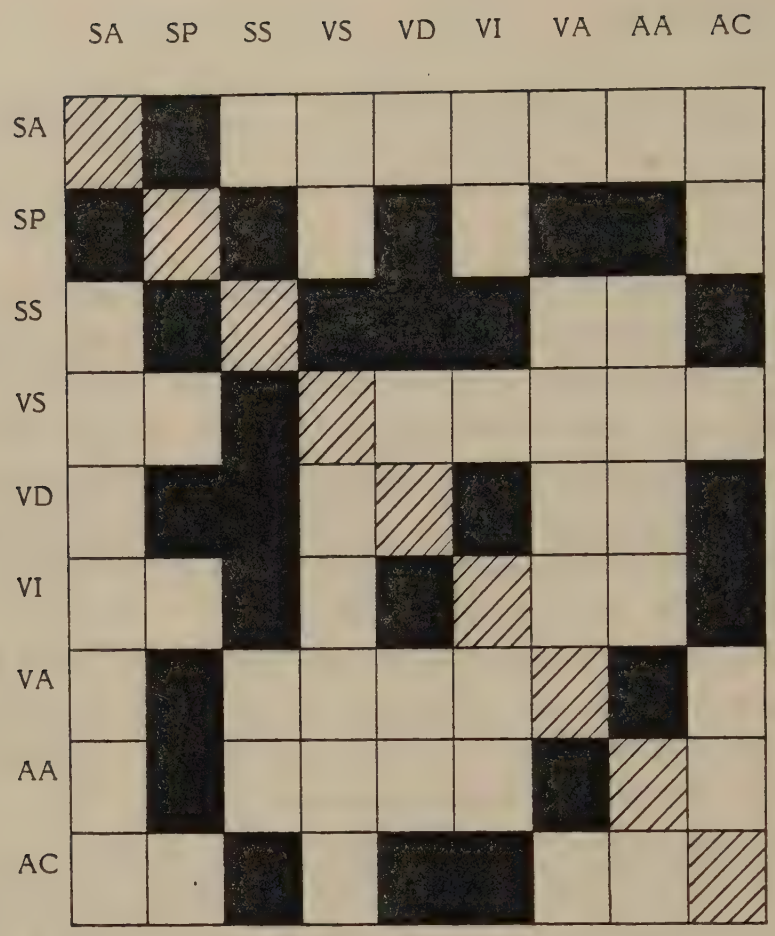
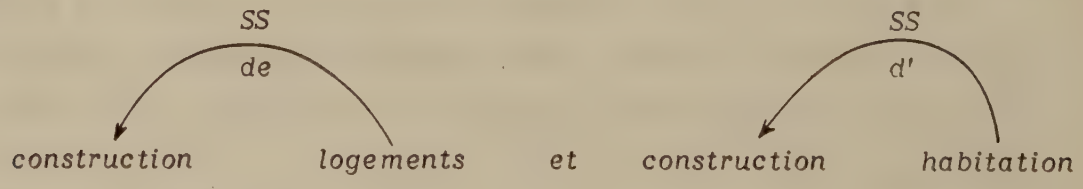


fig. 3. Matrice des synonymies lexico - syntaxiques potentielles

Remarquons que la diagonale principale (cases hachurées) n'implique pas uniquement l'identité des RLS : une synonymie doit pouvoir être établie entre par exemple :



Les mêmes équivalences peuvent être représentées par le graphe :

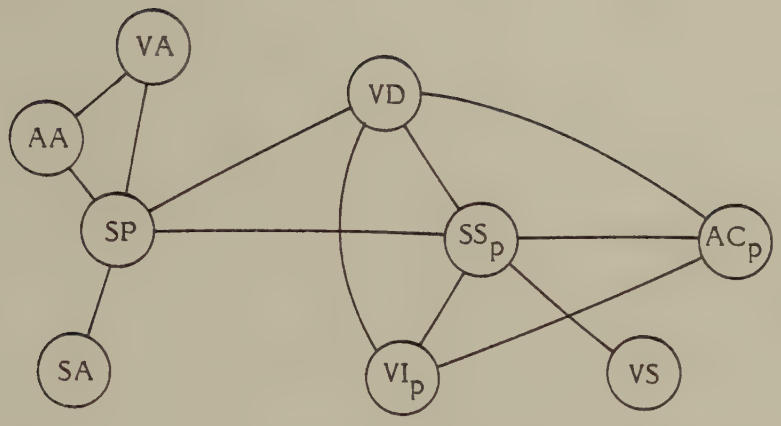


fig. 4. Synonymies lexico-syntaxiques potentielles

Nota : (XX) est implicite.

Remarque sur la non-transitivité :

Ce graphe laisse apparaître la non-transitivité des équivalences potentielles.

Prenons l'exemple "pas d'équivalence AA ~ SS_p" malgré "AA ~ SP" et "SP ~ SS_p". Il peut être illustré comme suit :

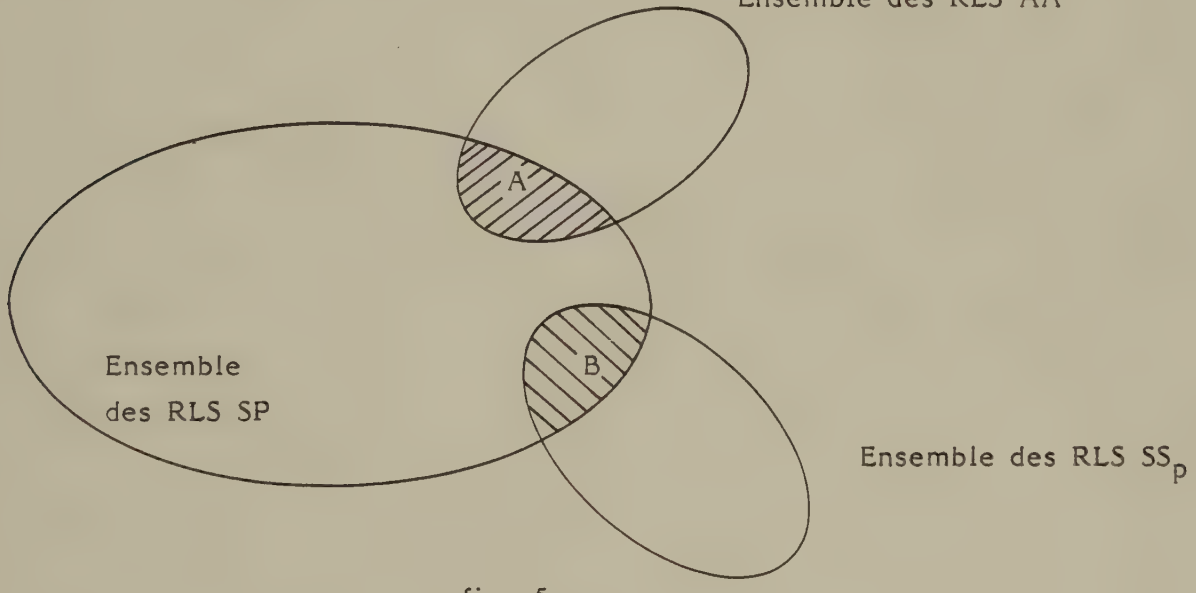


fig. 5.

On observe que : $A \cap B = \emptyset$

c'est à dire qu'il n'y a pas de RLS SP qui possède à la fois un équivalent parmi les RLS AA et un équivalent parmi les RLS SS_p.

V. Conclusion :

En guise de conclusion nous rappellerons que les RLS ainsi sélectionnées correspondent aux relations de la grammaire traditionnelle les plus répandues

Cet ensemble n'est pas exhaustif. Aussi le système devra-t-il présenter les caractéristiques d'un système ouvert : possiblité de définir et d'intégrer d'autres RLS. A cet égard la figure 1 peut suggérer d'autres relations.

Enfin, on peut noter qu'eu égard à leur aspect général, les transformations envisagées ici peuvent être rapprochées du *"système de périphrases"* de A.K. ZHOLKOVSKIJ et I.A. MEL'CHUK : *"Dans les divers travaux "transformationnels" on considère les transformations comme des substitutions équivalentes dans lesquelles les unités lexicales (lexèmes) restent inchangées ; seuls changent les morphèmes grammaticaux (terminaisons, mots outils, etc.), c'est à dire les constructions syntaxiques. Par contre, notre système de périphrases n'est pas limité à la synonymie syntaxique, mais est plutôt orienté vers la synonymie lexicale au sens le plus large."* (115, p.34) .

Chapitre troisième

COMPOSANTE SYNTAXIQUE DU SYSTEME

DESCRIPTION SYNTAGMATIQUE DE LA PHRASE

Introduction :

Après avoir soulevé, dans le premier chapitre, les problèmes que posait sous l'angle sémantique, la reconnaissance des relations de dépendance, nous avons, dans le deuxième chapitre, décrit l'allure que pourrait prendre la composante sémantique nécessaire pour résoudre les ambiguïtés inhérentes à cette reconnaissance.

Dans ce troisième chapitre, nous allons décrire la composante syntaxique qui nous permettra d'acquérir les éléments de base de la composante sémantique du système, à savoir les relations LS syntagmatiques.

Certains traitements qui en relèvent ne seront cependant pas abordés dans le cadre du présent exposé(*). Il s'agit de l'analyse grammaticale dont le but assigné ici est de résoudre les ambiguïtés dues à l'homographie grammaticale ("car" est conjonction ou substantif ; "outre" est substantif ou adverbe ; "contre" est préposition, verbe ou encore substantif ; "le" est article ou pronom ; etc.).

A fortiori, l'analyse morphologique (découpage ; attribution d'informations grammaticales aux unités lexicales ; etc.) effectuée dans une étape antérieure à cette analyse grammaticale, ne sera pas non plus abordée.

Rappelons néanmoins que cette analyse morpho-grammaticale a :
- pour entrée : une chaîne de caractères constituant le texte (celui-ci n'ayant subi aucun prétraitement manuel), et,
- pour sortie : une structure dans laquelle les unités lexicales ont été reconnues de même que les informations grammaticales qui leur sont associées dans le texte (catégories grammaticales, genre-nombre, etc.) (cf. fig. 6).

A l'issue des traitements morpho-grammaticaux, un texte T :

$$T = c_1\ c_2\ \ldots\ c_n$$

peut donc être représenté comme suit :

$$T = m_1\ m_2\ \ldots\ m_n$$

$$V = v_1\ v_2\ \ldots\ v_n$$

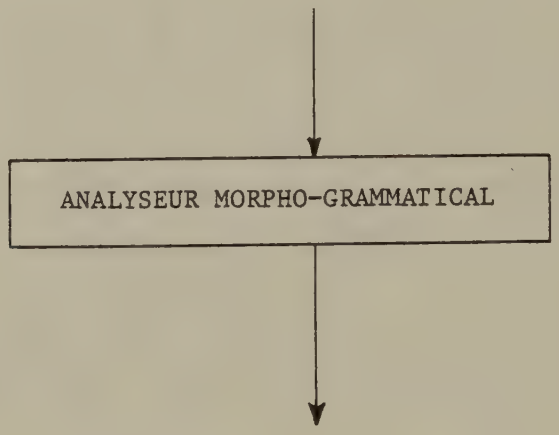
$$W = w_1\ w_2\ \ldots\ w_n$$

(*) On pourra se rapporter à cet effet aux thèses et publications suivantes, où ils sont très largement décrits : 2, 3, 5, 46, 60 et 68.

(1)

PHRASE NO :104

LES CONSTRUCTIONS DE CARACT@ERE EXP#ERIMENTAL PEUVENT -ETRE EXEMPT#EES DU PERMIS DE CONSTRUIRE PAR ARR-ET#E DU MINISTRE DE LA RECONSTRUCTION ET DE L' URBANISME , PRIS APR@ES AVIS MOTIV#E DU MAIRE .



(2)

N°	Mot	Valeur grammaticale	Genre-Nombre MS MP FS FP
3479	LES	ARTD	0101
3480	CONSTRUCTIONS	S	0001
3481	DE	PREPART	1111
3482	CARACT@ERE	S	1000
3483	EXP#ERIMENTAL	ADJP	1000
3484	PEUVENT	INDC	0101
3485	-ETRE	INF ETAT	1111
3486	EXEMPT#EES	PPAS AT	0001
3487	DU	PREPART	1000
3488	PERMIS	S	1100
3489	DE	PREP INF	0000
3490	CONSTRUIRE	INF	1111
3491	PAR	PREPART	0000
3492	ARR-ET#E	S	1000
3493	DU	PREPART	1000
3494	MINISTRE	S	1000
3495	DE	PREP	1111
3496	LA	ARTD	0010
3497	RECONSTRUCTION	S	0010
3498	ET	COJC REF	1111
3499	DE	PREP	1111
3500	L'	ARTD	1010
3501	URBANISME	S	1000
3502	,	,	0000
3503	PRIS	PPAS ADJP	1100
3504	APR@ES	PREPART	1111
3505	AVIS	S	1100
3506	MOTIV#E	PPAS ADJP	1000
3507	DU	PREPART	1000
3508	MAIRE	S	1000
3509	.	.	0000

fig. 6. A la chaîne d'entrée (1) correspond en sortie la structure (2). Celle-ci constitue l'entrée du système que nous présentons.

Nota : Les codes d'accents sont indiqués en annexe.

où m_1 est le mot de rang i du texte, v_i la valeur grammaticale correspondante dans le texte et w_i la valeur du genre-nombre associé à ce même mot m_i.

C'est précisément cette structure sous forme de triplets qui constitue l'entrée du système que nous présentons.

Ceci peut être représenté schématiquement comme sur la figure 7.

Notons que la frontière (trait discontinu) entre traitements morpho-grammaticaux et traitements syntaxico-sémantiques n'est pas imperméable. Il peut subsister en effet des ambiguïtés grammaticales qui demandent, pour leur résolution, un information en retour (flèche en pointillé, fig.7) provenant des traitements de niveau supérieur, syntaxico-sémantiques en l'occurrence (cf. ch. III, fig. 48(b) et fig. 67(b)).

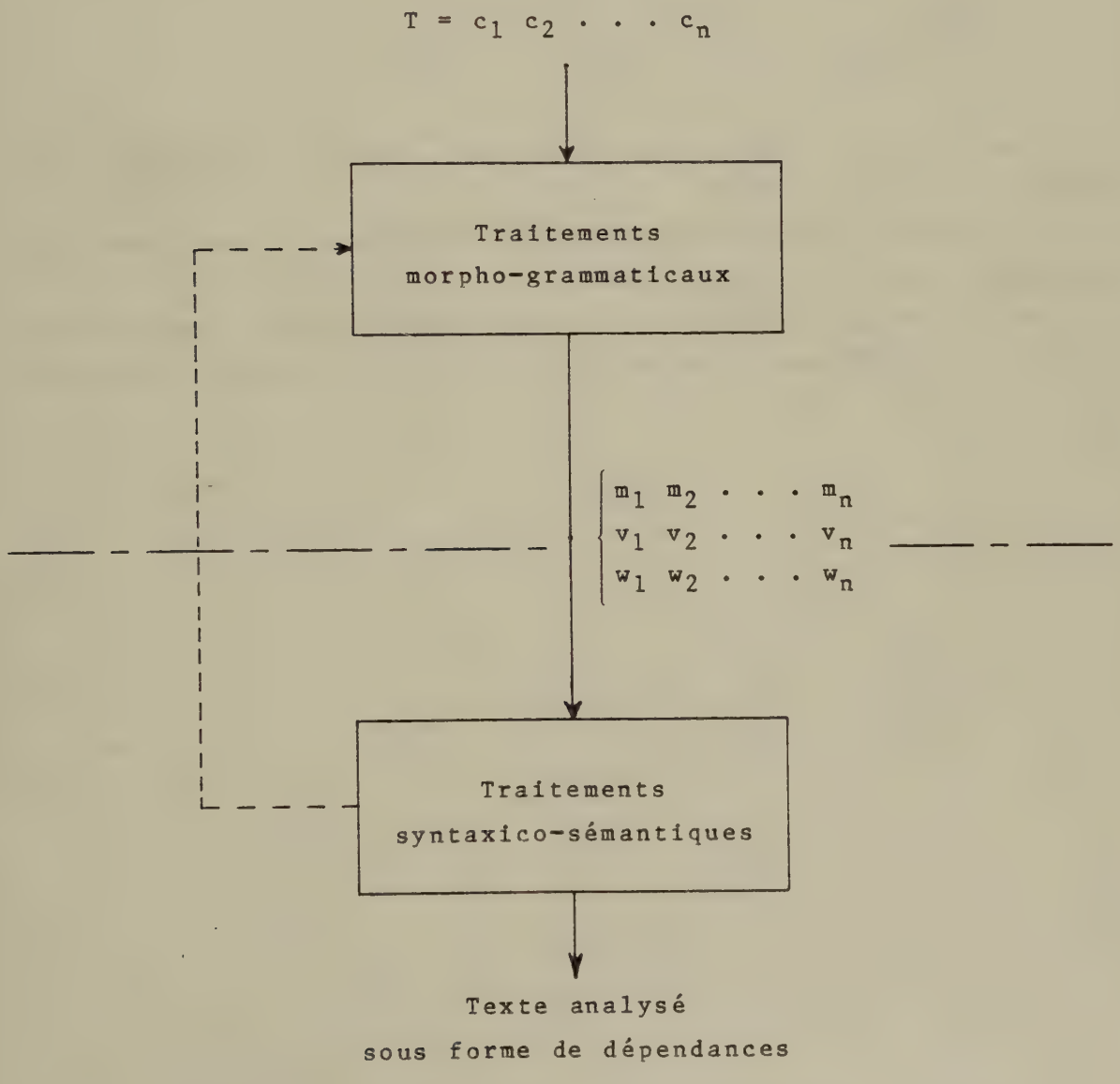


fig. 7.

Pour illustrer ce dernier point, donnons un exemple. A cet effet, considérons les trois phrases suivantes :

(1) *Des arbres poussent dans la forêt*
ARTI S INDC PREP ARTD S
PREPART

(2) *Des arbres peuplent les forêts*
ARTI S INDC ARTD S
PREPART

(3) *Des arbres partent les oiseaux*
ARTI S INDC ARTD S
PREPART

Dans (1), (2) et (3) l'ambiguïté sur "des"(*) ne peut être levée par des critères purement positionnels.

Une analyse syntaxique globale est nécessaire dans (1) pour lever cette ambiguïté : obligation d'avoir un sujet.

Par contre dans (2) et (3) qui sont de même structure, cette même ambiguïté doit (de préférence) être résolue par une analyse syntaxico-sémantique.

(*) La bonne valeur est soulignée.

première partie

SEGMENTATION DE LA PHRASE

I. Notion de filtre•:

Ce paragraphe vise à décrire de façon quasi-exhaustive l'outil que nous appelons filtre. Dans une lecture rapide du présent travail il n'est pas nécessaire de s'y attarder.

Soient :

M_L : L'ensemble des mots associés à une langue L ;

V_L : un ensemble de catégories grammaticales définies pour L ;

W_L : l'ensemble des catégories morphologiques définies pour L (ce sera, pour le Français, les catégories du genre-nombre).

Nous noterons E le sous-ensemble de base défini comme suit :

$E = \{ (m_i, v_i, w_i) \mid m_i \in M_L, v_i \in V(m_i)^{(*)}, w_i \in W(m_i)^{(*)} \}^{(**)}$

$\subset M_L \times V_L \times W_L$

Soit T un texte dont l'analyse morpho-grammaticale a été effectuée. Celui-ci est donc un élément de E^*.

$T = (m_1, v_1, w_1)\ (m_2, v_2, w_2) \ldots (m_n, v_n, w_n)$

En notant e un élément de E, nous écrirons également :

$T = e_1\ e_2 \ldots e_n$

I. 1. Filtre élémentaire :

I. 1. 1. Définition :

Nous appellerons filtre élémentaire tout sous-ensemble, noté F, de l'ensemble de base E.

$F = \{ (m_1, v_1, w_1), \ldots, (m_p, v_p, w_p) \}$

I. 1. 2. Cas particuliers :

- Si $F = \emptyset$ nous dirons du filtre qu'il est "passe - rien"
- Si F = E nous dirons que le filtre est "passe - tout"
- Si card F = 1 le filtre sera dit "unitaire"

  Exemple : $F = \{ (le, ARTD, MS) \}$

(*) $V(m_i)$ est l'ensemble des valeurs grammaticales hors contexte du mot m_i et $W(m_i)$ celui de ses valeurs genre-nombre.

(**) Nous utiliserons dans tout ce qui suit les symboles < et > pour représenter les accolades

• Si F est tel que :

$F = <(m, y, z)>$ (rappelons que $y \in V(m)$ et $z \in W(m)$)

où m est une unité lexicale donnée, nous dirons alors que le filtre est lexical Nous le noterons F_l^m.

Exemple : Le filtre lexical déteminé par l'unité "le" sera :

$F_l^{le} = <(le, ARTD, MS), (le, PPV, MS)>$

Notation : Nous écrirons plus simplement :

$F_l^m = <(m, *, *)>$

Plus généralement, si $M = <m_1, m_2, \ldots, m_p>$ est un ensemble de mots donnés, le filtre lexical défini parM sera :

$F_l^M = <(m_i, y, z) | m_i \in M, y \in V(m_i), z \in W(m_i)>$

Il est clair que :

$$F_l^M = \bigsqcup_{m_i \in M} F_l^{m_i}$$

En utilisant la notation "étoilée" vue plus haut :

$F_l^M = <(m_i, *, *) \mid m_i \in M>$

Exemple : Si $M = <à, au, aux>$ alors :

$F_l^M = <(à, *, *)> \cup <(au, *, *)> \cup <(aux, *, *)>$

• De même, si F est tel que :

$F = <(x, v, z) \mid x \in M_L$ et $v \in V(x), z \in W(x)>$

où v est une valeur grammaticale donnée, nous dirons alors que le filtre est "grammatical" et qu'il est associé à la valeur v. Il sera noté F_g^v.

Exemple : Le filtre grammatical déterminé par ARTI sera :

$F_g^{ARTI} = <(*, ARTI, *)>$

En développant cette expression nous aurons :

$F_g^{ARTI} = <(un, ARTI, MS), (une, ARTI, FS), (des, ARTI, MP), (des, ARTI, FP)>$

Plus généralement, si $G = \langle v_1, v_2, \ldots, v_p \rangle$ est un ensemble de valeurs grammaticales données, le filtre grammatical défini par G sera :

$F_g^G = \langle (x, v_i, z) \mid x \in M_L \text{ et } v_i \in V(x), z \in W(x) \rangle$

Il est clair que :

$$F_g^G = \bigcup_{v_i \in G} F_g^{v_i}$$

Exemple : Le filtre grammatical associé à $G = \langle ARTD, ARTI \rangle$ est :

$F_g^G = \langle (*, ARTD, *) \rangle \cup \langle (*, ARTI, *) \rangle$.

• Si F est tel que :

$F = \langle (m, v, z) \mid z \in W(m) \rangle$

où m est une unité lexicale donnée et v sa valeur grammaticale, nous dirons alors que le filtre est "lexico-grammatical" ou encore "mixte" et nous le noterons F_x.

Exemple : $F_x = \langle (physique, SUBS, *) \rangle$

Développée, cette expression donnera :

$F_x = \langle (physique, SUBS, MS), (physique, SUBS, FS) \rangle$

Comme précédemment, on peut associer le filtre lexico-grammatical à un ensemble de mots accompagnés de leurs valeurs grammaticales.

• Si enfin, F est tel que :

$F = \langle (x, v, w) \mid x \in M_L \text{ et } v \in V(x), w \in W(x) \rangle$

où v et w sont donnés, nous dirons alors que le filtre est "morpho-grammatical".

Exemple : $F = \langle (*, SUBS, MS) \rangle$

Nous n'avons pas jugé utile de retenir le dernier cas théoriquement possible où seul w serait donné à l'avance.

• **Filtre complémentaire :** C'est le filtre noté $\bar{F}$ (F barre) qui correspond au complément, dans E, de l'ensemble des triplets définissant F :

$$\bar{F} = E - F$$

Nous l'appellerons aussi "filtre passe-tout sauf"

I. 2. Filtrage :

Soit F un filtre élémentaire et T un texte.

Nous appellerons filtrage l'opération qui consiste, en appliquant F à T, à extraire toutes les occurrences des éléments de F dans T.

Ex. 1 : Considérons le filtre F = < le, ARTD, MS >

L'application de ce filtre à un texte T en extraira toutes les occurrences de l'unité lexicale "le" ayant pour valeur grammaticale "article défini" et pour valeur genre-nombre "masculin-singulier".

Ex. 2 : Considérons le filtre lexico-grammatical :

F_x = < (à, PREP, *) > ∪ < (à, PREPART, *) >

Ce filtre extraira d'un texte T quelconque toutes les occurrences de l'unité lexicale "à" en tant que *préposition ou préposition article* et non en tant que *préposition introduisant un infinitif* .

Ex. 3 : Soit le filtre lexical :

F_1 = < (de, *, *) > ∪ < (du, *, *) > ∪ < (des, *, *) > ∪ < (d', *, *) >

Appliqué à un texte T ce filtre en extraira toutes les occurrences des unités lexicales *"de","du","des" ou "d'"* quelles que soient leurs valeurs grammaticales ou genres-nombres dans le texte.

Ex. 4 : Considérons le filtre grammatical :

F_g = < (*, ADJA, *) > ∪ < (*, ADJP, *) >

Ce filtre extraira d'un texte T quelconque toute occurrence d'un *adjectif antéposé ou postposé* .

Ex. 5 : Considérons le filtre F = < (. , *, *) >

L'application du filtre complémentaire $\bar{F}$ à un texte T quelconque extraira les occurrences de tout mot différent du point.

I. 3. Filtre syntaxique :

I. 3. 1. Concaténation de filtres élémentaires :

Soit F_1 et F_2 deux filtres élémentaires.

La concaténation de F_1 et F_2, notée F_1F_2, définit un ensemble de configurations d'ordre 2 obtenues en formant le produit cartésien $F_1 \times F_2$ puis en concaténant les composantes dans chaque couple (GROSS & LENTIN, 65)

$$F_1F_2 = \langle xy \mid x \in F_1, y \in F_2 \rangle$$

I. 3. 2. Définition :

Nous appellerons filtre syntaxique d'ordre p l'ensemble des configurations d'ordre p défini par la concaténation de p filtres élémentaires :

$$\Phi^p = F_1F_2 \ldots F_p$$

Plus généralement, nous appellerons filtre syntaxique toute réunion de filtres syntaxiques d'ordre quelconque.

$\Phi = \Phi^{P_1}_1 \quad \Phi^{P_2}_1 \quad \dots \quad \Phi^{P_n}_n$

NB. : Un filtre syntaxique est un élément de $\mathfrak{P}(E^*)$, c'est à dire un ensemble de textes.

I. 4. Filtrage syntaxique :

Soit P un filtre syntaxique et T un texte.

Nous appellerons filtrage syntaxique l'opération qui consiste, en appliquant P à T, à extraire toutes les occurrences des éléments de P dans T.

Exemple : Soit les filtres élémentaires :

$F_1 = <(*, \text{ARTD}, *)>$;

$F_2 = <(*, \text{SUBS}, *)>$;

$F_3 = <(\text{de}, \text{PREPART}, *)> \cup <(\text{d'}, \text{PREPART}, *)>$

Le filtre syntaxique $P = F_1F_2F_3F_2$ appliqué à un texte T en extraira toutes les occurrences des configurations appartenant ou définies par P. En particulier, si l'on a :

T = La physique des solides ...

ARTD S PREPART S V -

alors P extraira :

La physique des solides

ARTD S PREPART S

FS FS MFP MP

Remarque :

Les filtres tels que nous les entendons ici, bien que procédant de la notion d'ensemble, ne correspondent pas aux filtres définis en topologie (BOURBAKI, TG I. 6, § 6).

I. 5. Extension de la notion de filtre :

La notion de filtre a été introduite à partir d'un filtre unitaire composé de trois éléments ou champs d'information déterminés.

L'application d'un filtre syntaxique réalisait la simple extraction des configurations correspondant au filtre, sans transformation de celles-ci ou plus généralement du texte, entrée du filtre.

En fait cette notion peut être étendue dans deux directions :

a) Le nombre de champs d'information peut être variable selon le type de filtre. On peut envisager d'ajouter un (ou plusieurs) champ(s) d'information ou d'en supprimer. Le filtre unitaire de base serait non plus un triplet mais un n-uplet.
Cette possibilité a été entrevue lorsque nous avons parlé des filtres lexicaux grammaticaux ou encore mixtes.

b) L'action des filtres peut s'accompagner d'une transformation dans le texte des configurations sélectionnées. Cette transformation peut avoir pour effet de modifier d'une part le nombre des champs d'information associés aux mots dans le texte et d'autre part de modifier la valeur d'un champ donné. Dès lors, un texte pourra être représenté comme une séquence de n-uplets dont les éléments ou champs constitutifs seront ou non pris en compte, selon les buts recherchés. On écrira :

$$T = \begin{cases} m_1\ m_2 \ldots m_n \\ v_1\ v_2 \ldots v_n \\ \\ z_1\ z_2 \ldots z_n \end{cases}$$

La notion de filtre ainsi introduite définit l'outil de base qui nous permettra de reconnaître les RLS syntagmatiques.

II. Resserrement du champ d'investigation :

Comme nous l'avons vu, les relations LS revêtent souvent des configurations où les termes mis en relation sont non contigus. C'est d'ailleurs cet aspect qui constitue une des sources des problèmes posés (*).

II. 1. - Dans un premier temps la recherche et la construction des diverses relations de dépendance sous-tendues par le discours se feront à l'intérieur d'une même unité, à savoir la phrase prise au sens traditionnel du terme.

La ponctuation forte, c'est à dire le point, le point d'exclamation, le point d'interrogation et les points de suspension constitueront donc dans un premier temps des "barrières" qui ne pourront pas être franchies. Nous entendons par là que les termes situés de part et d'autre d'une "barrière" ne pourront être mis en relation.

(*) Il y a aussi la polyvalence des prépositions, par exemple : *"il a été renversé par un beau matin d'avril"* ; et d'autres.

Il convient cependant de relever qu'une telle option conduit à certains sacrifices, dont l'essentiel, du point de vue où l'on se place, réside dans la non-prise en compte des RLS sous-jacentes ou impliquées par les anaphores, ce dernier terme étant pris dans un sens très large.

Ainsi, dans :

(1) *Aristote, né à Stagire en 384 av. J.C. fut l'un des plus grands savants de l'Antiquité.*

(2) Il fonda à Athènes sa propre école philosophique.

(Histoire de la Logique, A. MAKOVELSKI ; p. 115)

la non-détermination de la connexion anaphorique qui lie "Il" dans (2) à "Aristote" dans (1) conduit à ne pas répertorier la relation LS qui s'établit pourtant entre "Aristote" et "fonder".

Nous verrons par contre (cf. ch. V, § III & IV) que si les liens d'ordre sémantique qui s'établissent entre phrases différentes sont de la sorte exclus (e.g. liens anaphoriques), lors de cette étape, d'autres liens, également sémantiques, pourront néanmoins être établis lors de l'étape d'analyse sémantique, de sorte que, si l'analyse préconisé ne déborde pas le cadre de la phrase sur le plan syntaxique, elle le déborde sur le plan sémantique.

II. 2. - Une deuxième limitation du champ d'investigation virtuelle cette fois, car simplificatrice mais non rigide, sera basée sur le fait que les relations LS revêtent des formes superficielles qui emprisonnent souvent les éléments de la relation dans une même chaîne (nominale ou verbale). Nous avons appelé homosyntagmatiques ces relations et hétérosyntagmatiques celles qui s'établissent entre éléments appartenant à des chaînes différentes (cf. ch. II, § III. 1.).

La segmentation de la phrase en chaînes nominales et/ou verbales peut donc conduire à un autre type de barrières ou séparations qui permettront de resserrer encore davantage le champ d'investigation ou de recherche des RLS.

Ces barrières pourront cependant, contrairement aux premières, être franchies (relations hétérosyntagmatiques), mais une option limitative accompagnera cette possibilité, du moins dans le cadre du présent exposé. Seules, en effet, seront prises en compte les relations hétérosyntagmatiques qui s'établissent entre éléments appartenant à des chaînes de nature différente.

Autrement dit, les seules RLS hétérosyntagmatiques qui seront prises en compte seront celles qui s'établissent entre un élément d'une chaîne nominale et un élément d'une chaîne verbale.

Cette option, souvent justifiée, conduit malgré tout à certains sacrifices. Ainsi, dans la phrase suivante :

/ Le conseil supérieur / élit / son vice-président / parmi ses membres /
CN CV CN CN
(Règlementation de l'urbanisme)

découpée comme indiqué (les barrières sont représentées par le symbole "/"), aucune relation LS ne sera établie entre éléments appartenant à des chaînes nominales différentes, et, par suite, la relation LS qui pourrait, par exemple, être établie entre "membres" et "conseil" ("membres du conseil") ne sera pas déterminée.

Il convient de remarquer que la construction de cette relation suppose au préalable la reconnaissance de la connexion anaphorique qui s'établit entre "ses" et "conseil".

Les mêmes remarques peuvent être formulées pour "vice-président" et "conseil".

II. 3. - Signalons enfin, dès maintenant, une troisième limitation du champ d'investigation qui sera généralement spécifique à la recherche d'une (ou plusieurs) RLS donnée(s).

Cette troisième limitation sera en effet fondée sur la reconnaissance d'une configuration fixée à l'avance, lors de la recherche d'une RLS donnée. Cette configuration, reconnue, constitue alors une barrière ou frontière au delà de laquelle toute recherche serait vaine, ou encore, conduirait à des erreurs dont on ne peut s'accommoder à l'étape du traitement considérée. Mais comme la configuration aura elle-même été établie par une heuristique, il apparaîtra des erreurs dues, comme il est habituel, à l'application normale de cette heuristique (cf. ch. III, 2^{e} Partie, fig. 55 (b)).

III. Segmentation de la phrase :

La segmentation de la phrase est basée sur la reconnaissance des différentes chaînes nominales et/ou verbales qui y sont incluses.

La phrase, que nous désignerons par le symbole P, peut dès lors être considérée comme une succession de chaînes nominales et/ou verbales, éventuellement séparées par des mots ou séquences de mots permettant leur articulation.

Ces mots ou séquences de mots, que nous désignerons par le symbole X, sont à la base des compositions ou enchâssements de ce qu'il est convenu d'appeler phrases simples. Ce sont les conjonctions de coordination, les conjonctions de subordination, les pronoms relatifs ou encore la ponctuation.

Exemples :

/ Ces dispositions / ne sont pas opposables / aux personnes / qui /
CN_1 CV_1 CN_2 X
auront déposé / une demande de permis de construire /.
CV_2 CN_3

/ Les faces directement accessibles des conduits adossés à un mur extérieur /
CN_1
doivent avoir / une isolation suffisante / pour que / le refroidissement/
CV_1 CN_2 X CN_3
ne contrarie pas / le tirage /.
CV_2 CN_4

/ Ces sections / , / de caractère permanent / , / doivent comprendre /
CN_1 X CN_2 X CV_1
douze membres / au minimum / et / vingt membres / au maximum /
CN_3 CN_4 X CN_5 CN_6
(Règlementation de l'urbanisme)

Remarque :

La segmentation proposée ici reste différente de la segmentation qui est effectuée dans les analyses de la phrase en propositions (82) . Il est clair cependant qu'elle peut être considérée comme un étape préliminaire pour une segmentation en propositions.

Les hypothèses de travail restent par contre sensiblement les mêmes, à savoir qu'il est possible, dans la grande majorité des cas, d'identifier :
- dans un cas *"le début et la fin de la proposition française à partir d'un petit nombre de marques ou de formes"* (82, p. 14) ;
- dans notre cas : le début et la fin des chaînes nominales ou verbales, et ce, au moyen de filtres mettant en oeuvre un petit nombre de composants et dont nous allons donner maintenant la description.

III. 1. Reconnaissance des chaînes nominales :

III. 1. 1. Caractéristiques des chaînes nominales :

Du point de vue structural, les chaînes nominales peuvent être caractérisées d'une part, par la nature des éléments qui les constituent, d'autre part, par les règles d'agencement de ces éléments.

En effet, dans leur quasi-totalité, les chaînes nominales sont construites à partir des seules catégories grammaticales suivantes :

- Les articles (définis, indéfinis, partitifs)
- Les substantifs
- Les adjectifs (déterminatifs, qualificatifs)
- Les verbes à l'infinitif ou au participe passé
- Les adverbes
- Les prépositions
- Les conjonctions.

Cependant, tout agencement de ces catégories ne constitue pas forcément une chaîne nominale, quand bien même cet agencement serait issu d'un texte.

Ainsi, dans la phrase suivante :

"J'appelle un chat un chat, et Rolet un fripon"
(BOILEAU)

le segment "un chat un chat" ne représente pas une chaîne nominale malgré la conformité des éléments dont il est formé.

Le filtre syntaxique de reconnaissance des chaînes nominales, que nous noterons P_N, devra par conséquent tenir compte :

- et de la nature des éléments constitutifs de la CN,
- et de la syntaxe de ces éléments.

III. 1. 2. Matrice de reconnaissance des chaînes nominales :

Un filtre est un ensemble de configurations . En l'occurrence, P_N sera défini par l'ensemble des configurations nominales.

Il est clair cependant que l'on ne peut définir P_N en extension. Il nous faut donc le définir en compréhension.

Pour cela, nous faisons appel à une matrice dont les lignes et les colonnes sont des filtres élémentaires ayant pour but la reconnaissance des éléments constitutifs des chaînes nominales, le corps de la matrice représentant la syntaxe de ces éléments. Nous retrouvons ainsi les deux caractéristiques des chaînes nominales vues plus haut.

III. 1. 2. 1. Lignes et colonnes de la matrice :

Chaque ligne et chaque colonne de la matrice contiendra une information relative à un filtre élémentaire propre à reconnaître un composant spécifique de la chaîne nominale.

Une liste générale de ces composants a été donnée plus haut. Nous énumérons ci-après, à titre indicatif, la liste des filtres élémentaires qui ont été retenus.

a) Les articles : leur reconnaissance est obtenue au moyen du filtre élémentaire grammatical suivant :

$f_g^{<Art>} = <$ *articles définis, articles indéfinis, articles déterminatifs* $>$

b) Les substantifs :

$f_g^{<S>} = <$ *Substantifs* $>$

c) Les adjectifs : selon qu'ils sont antérieurs ou postérieurs, ils sont reconnus par :

$f_g^{<AdjA>} = <$ *adjectifs antérieurs* $>$

$f_g^{<AdjP>} = <$ *adjectifs postérieurs, verbes au participes passé en position d'épithète* $>$

d) Les verbes :

$f_g^{<Inf>} = <$ *verbes à l'infinitif* $>$

e) Les adverbes :

$f_g^{<Adv>} = <$ *adverbes modifieurs des adjectifs et des participes passés épithètes* $>$

f) Les prépositions :

$f_g^{<Prep>} = <$ *prépositions, prépositions-articles, prépositions introduisant un infinitif* $>$

g) Les conjonctions :

$f_g^{<Conj>} = <ET,\ OU>$

Remarques :

- Cet ensemble de filtres élémentaires, bien que n'étant pas exhaustif, reflète néanmoins la composition de la plupart des chaînes nominales des corpus traités. Notons toutefois que les méthodes utilisées rendent aisée toute extension ou modification de ces filtres.
- Toute définition d'un filtre élémentaire nouveau revient en fait à définir une nouvelle classe ou catégorie de mots.

Ainsi, dans la reconnaissance des chaînes nominales les articles définis, les articles indéfinis, mais aussi les adjectifs déterminatifs sont réunis dans une même classe. Il en est de même des prépositions qui, à ce stade reçoivent le même statut, mais qui à d'autres étapes du traitement sont réparties dans des classes différentes, voire individualisées (cf. ch. III, § IV. 2. 1. 1.).

Nous rencontrons là une des caractéristiques fondamentales du sytème, à savoir la possibilité de définir à tout moment les classes jugées nécessaires pour une étape donnée du traitement.

Rappelons enfin que l'action d'un filtre élémentaire peut s'accompagner d'une attribution, aux unités détectées, de la catégorie nouvellement définie.

III. 1. 2. 2. Corps de la matrice :

Le corps de la matrice est sensée représenter la syntaxe d'agencement (*) des différents éléments constitutifs de la chaîne nominale, et par conséquent des différents filtres élémentaires ayant pour objet la reconnaissance de ces éléments.

Par suite, toute succession de filtres élémentaires engendrée par la matrice est un filtre syntaxique apte à reconnaître un ensemble de chaînes nominales données.

Inversement, toute chaîne ou configuration appartenant à un texte est baptisée nominale si sa structure peut être engendrée par la matrice.

(*) Il s'agit des transitions possibles entre éléments constitutifs. La matrice est donc, pour nous, la représentation d'un automate d'état fini qui reconnait les chaînes, nominales dans le cas présent.

Remarque :

Un filtre élémentaire définissant en général une classe de mots, la représentation des chaînes nominales ainsi obtenue reflète en fait non pas l'ensemble des chaînes nominales réelles ou possibles, mais un ensemble plus vaste.

Cependant, ceci ne saurait être gênant dans le cadre que nous nous sommes fixés, où seule la reconnaissance est envisagée et non la génération.

L'élaboration de ladite matrice sera donc fondée sur les impératifs liés à la reconnaissance des diverses chaînes nominales, laquelle suppose :

- l'identification du début des chaînes en question,
- l'identification de leur fin,
- la conformité de leur syntaxe interne.

En effet, toute unité reconnue par les filtres élémentaires définis plus haut ne constitue pas systématiquement un début ou une fin possible pour un chaîne nominale.

Par exemple, on ne peut commencer une chaîne nominale par un adjectif postérieur ou encore par une conjonction de coordination. On ne peut non plus finir par un adjectif antérieur ou encore par un article.

Par ailleurs, comme nous l'avons vu, toute catégorie reconnue par un filtre élémentaire donné ne peut être suivie ou précédée de n'importe quelle autre catégorie reconnue par un autre filtre élémentaire.

Ainsi, dans la syntaxe interne de la chaîne nominale, un substantif ne peut être suivi d'un article alors que cette même succession est possible dans la langue (cf. exemple du paragraphe III. 1. 1.). Inversement, un article ne peut être précédé d'un substantif ou encore d'un adjectif postérieur.

Il convient de relever qu'il s'agit là de règles d'agencement, non pas orientées vers l'analyse grammaticale (la succession "substantif-article" est dans ce cas possible), mais plutôt vers la reconnaissance des chaînes nominales. Paradoxalement en effet, permettre la succession "article-conjonction", impossible dans la langue, serait sans effet sur la reconnaissance, puisque, a priori, une telle succession ne peut se rencontrer dans les textes analysés.

Ce sont ces considérations d'ordre positionnel qui président à l'élaboration de la matrice de reconnaissance des chaînes nominales.

a) Règles d'agencements des filtres élémentaires :

Pour matérialiser la possibilité qu'ont deux catégories ou filtres élémentaires donnés, f_i et f_j, de se succéder, on porte dans la matrice, à l'intersection de la ligne f_i et la colonne f_j, un nombre non nul (par exemple 1;

cases sombres sur les figures), la succession $f_i f_j$ non permise étant matérialisée par la présence d'un zéro.

Portant sur la succession de deux éléments, les règles d'agencement ainsi mises en jeu sont dites binaires.

Considérons à titre d'exemple la matrice suivante :

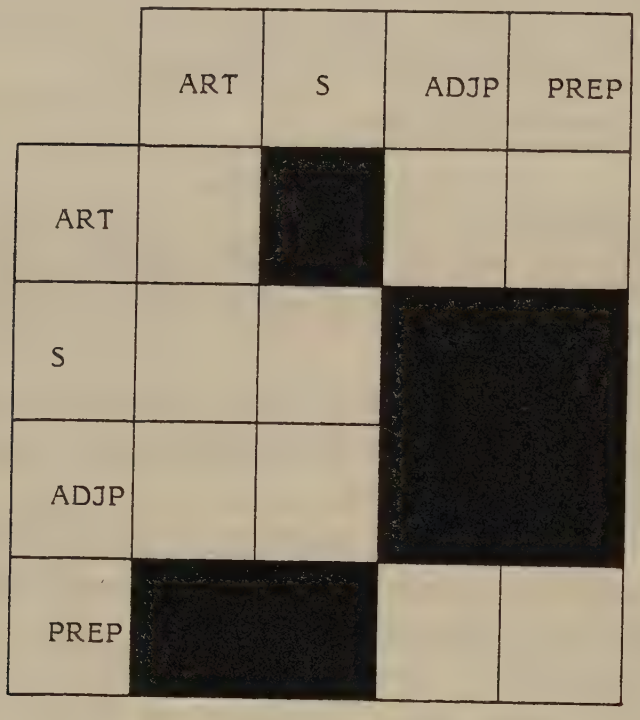


fig. 8.

qui équivaut au réseau de transition suivant :

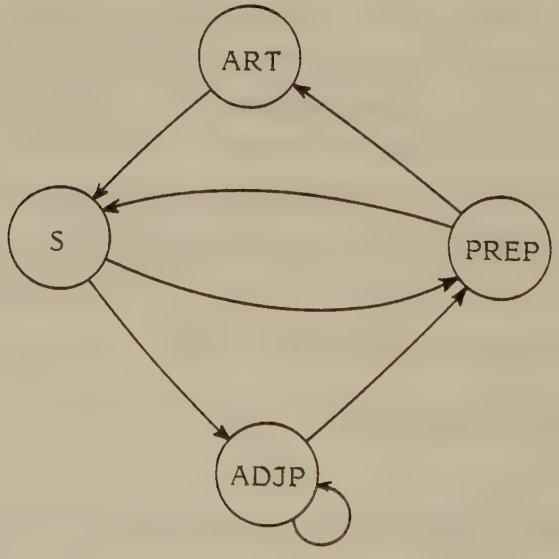


fig. 9.

Celle-ci renferme les règles d'agencement relatives aux filtres élémentaires permettant de reconnaître les catégories les plus répandues pour la formation des chaînes nominales (*articles, substantifs, adjectifs postérieurs, prépositions*).

Appliquée à un texte, cette matrice est apte à détecter les continuités syntaxiques indispensables pour la reconnaissance des chaînes nominales.

Inversement appliquée à une chaîne réputée nominale, elle est apte à détecter la continuité ou la discontinuité de la syntaxe interne de la chaîne proposée, et par conséquent de confirmer ou d'infirmer la proposition s'y rapportant.

Ainsi dans la phrase suivante :

"Galiléo Galilée construisit un téléscope agrandissant de trente fois les objets observés"

(Histoire de la logique, A. MAKOVELSKI, p. 389)

la configuration :

de trente fois les objets observés
PREP ART S ART S ADJP

ne sera pas reconnue comme formant une chaîne nominale puisque présentant une discontinuité. La succession S + ART n'est pas en effet autorisée.

Cette représentation de la syntaxe interne de la chaîne nominale est cependant incomplète. En effet, aucune contrainte relative aux éléments susceptibles de se trouver en début ou en fin de chaîne nominale n'est, pour le moment, présente dans la matrice.

b) Identification des début et fin des chaînes nominales :

L'identification des début et fin des chaînes nominales relève non seulement de la nature des éléments pouvant commencer ou finir une chaîne nominale, comme nous l'avons vu, mais aussi de leur syntaxe.

Il ne suffit pas en effet à une catégorie d'être un début ou une fin possible pour une chaîne nominale, pour qu'elle le soit à sa première occurrence. C'est par contre sa position relative vis à vis des autres catégories qui permet d'infirmer ou de confirmer définitivement son statut de début ou fin de chaîne nominale.

Ainsi, alors que la catégorie S peut être début, fin ou encore élément interne d'une chaîne nominale, que la catégorie ART peut être début mais aussi élément interne d'une chaîne nominale, dans la succession :

S ART

le premier élément annonce toujours la fin d'une chaîne nominale et le second le commencement d'une autre chaîne nominale.

Pour marquer la potentialité qu'ont certaines catégories d'être début ou fin de chaînes nominales, on associe à chaque ligne et à chaque colonne de la matrice un nombre tel que :

- lorsqu'il est différent de zéro, il indique la possibilité de la ligne (respectivement la colonne) d'être début (respectivement fin) d'une chaîne nominale,
- et, lorsqu'il est nul, l'impossibilité de la ligne ou de la colonne à avoir l'un ou l'autre de ces statuts.

Rappelons, en effet, que les conventions adoptées pour la construction de la matrice associent les lignes aux débuts des successions binaires (donc aux débuts des chaînes nominales) et les colonnes à leurs fins (donc à la fin des chaînes nominales).

La représentation schématique de ces informations se fait en hachurant les noms des lignes et colonnes concernées (cf. fig. 10). La matrice précédente devient :

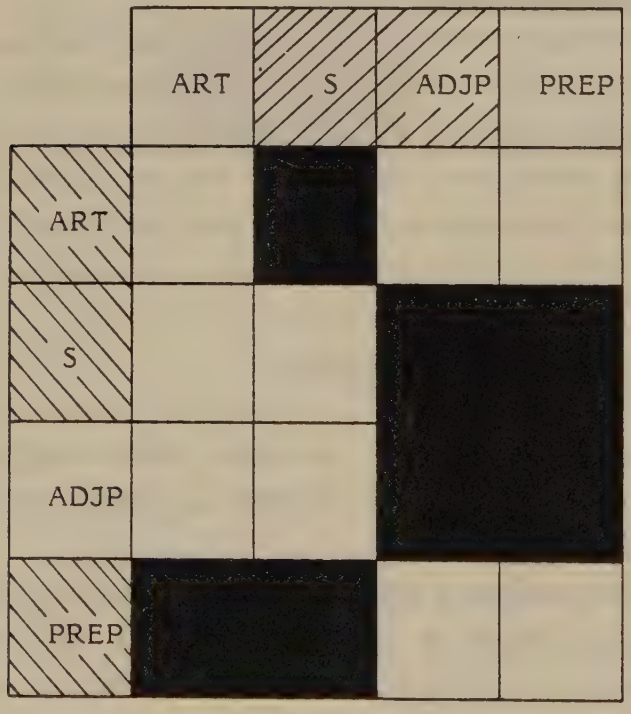


fig. 10.

que nous pouvons également représenter par le réseau de transition :

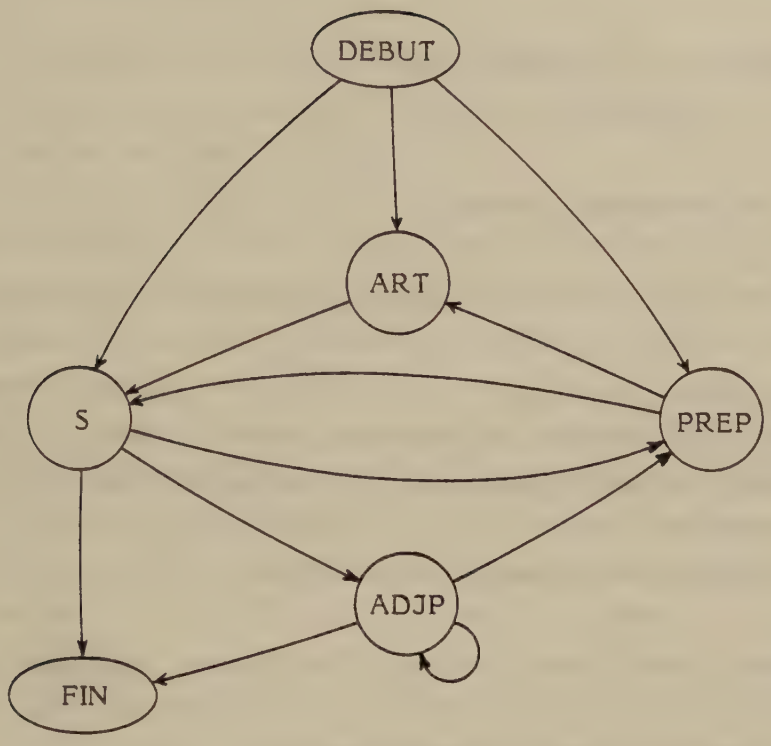


fig. 11

Remarques :

- Si la représentation proposée de la syntaxe interne des chaînes nominales reste relativement fruste, nous verrons que dans la pratique, elle permet de reconnaître les chaînes nominales avec des taux d'erreurs extrêmement faibles (cf. § III. 1. 3.)

Ces erreurs pourront d'ailleurs, souvent, être compensées dans les étapes ultérieures du traitement (cf. troisième type de segmentation ; utilisation de la connaissance). En effet, la détection de ces erreurs relève, presque toujours, non pas des règles d'agencement, mais de la composante sémantique du système, en l'occurrence les relations LS.

Ainsi, soumise à l'analyse, la phrase suivante :

P_1 : *"Les mathématiques conduisent l'homme du sensible à l'intelligible"*
(Histoire de la logique, A. MAKOVELSKI, p. 103)

laissera se dégager deux chaînes nominales dont une seule en fait, la première, est valide. Les continuités syntaxiques reconnues par la matrice précédente (cf. fig. 10) sont en effet :

(1) Les mathématiques
ART S
(2) l' homme du sensible à l' intelligible
ART S PREP S PREP ART S

La phrase P_1 est dès lors segmentée comme suit :

P_1 : *"/ Les mathématiques / conduisent / l'homme du sensible à l'intelligible /"*

Or, une telle segmentation ne peut être évitée à ce stade du traitement puisque l'on peut bâtir, à partir de la même charpente syntaxique, d'autres phrases admettant précisément cette segmentation, voire même des segmentations différentes de celle souhaitée.

Ainsi, alors que la véritable segmentation de P_1 est :

P_1:"/ Les mathématiques / conduisent / l'homme / du sensible / à l'intelligible / "

on trouve, bâties sur la même charpente syntaxique :

P_2:"/ Le ferrailleur / conduit / la voiture du client / à la casse / "
P_3:"/ Le programmeur / accepte / les remarques des initiés au Fortran / "
P_4:"/ Le président / reçoit / les cadeaux / des invités / à la chasse / "

- Toutefois et afin de minimiser encore davantage le nombre d'erreurs éventuelles, nous adopterons l'approximation selon laquelle les prépositions autres que < d', de, du, des, à, au, aux, en > (voir J.H. JAYEZ, 70) introduisent généralement, non pas des compléments de substantifs ou d'adjectifs, mais des compléments circonstanciels, cause principale des erreurs constatées. La préposition "par" sera cependant également retenue pour son emploi fréquent en tant qu'introducteur du sujet "vrai" dans les formes passives.

Cette option limitative, nécessaire dans une étape initiale, répond en fait à une attitude que nous retrouverons tout au long de ce chapitre. Compte tenu de l'enjeu, celle-ci consiste en effet à ne pas s'accommoder d'un excès d'erreurs pouvant se répercuter dans l'apprentissage des relations LS.

Autrement dit, nous préfèrerons nous accommoder d'un certain taux de non-reconnaissance ou de silence plutôt que d'emmagasiner une connaissance quelquefois erronée.

Cette attitude ne sera cependant maintenue que dans une première étape. Elle devra ensuite peu à peu disparaître, au fur et à mesure que s'accroîtra l'apport de la composante sémantique du système (cf. ch. V)

III. 1. 3. Expérimentation :

La figure 12 donne la matrice retenue pour l'expérimentation (*). Nous pouvons notamment y distinguer les filtres élémentaires (lignes et colonnes) qui ont été en définitive adoptés.

Ceux-ci sont précédés de leur type (G : grammatical, M : lexical, X : lexico-grammatical ou mixte) et suivis de l'information (0 ou 1) indiquant leur potentialité à être début (ligne) ou fin (colonne) de chaîne nominale.

Rappelons que cette matrice peut être modifiée à tout instant selon la qualité des filtrages obtenus.

Les figures 13 et 14 montrent sur quelques phrases, les unes issues d'un texte scientifique (cf. fig. 13), les autres de textes juridiques (cf. fig. 14), les reconnaissances effectuées à l'aide de cette matrice.

Les phrases 26 et 27 offrent notamment deux exemples d'erreurs dans la reconnaissance, alors que la phrase numéro un offre un cas de non-reconnaissance (ou de silence).

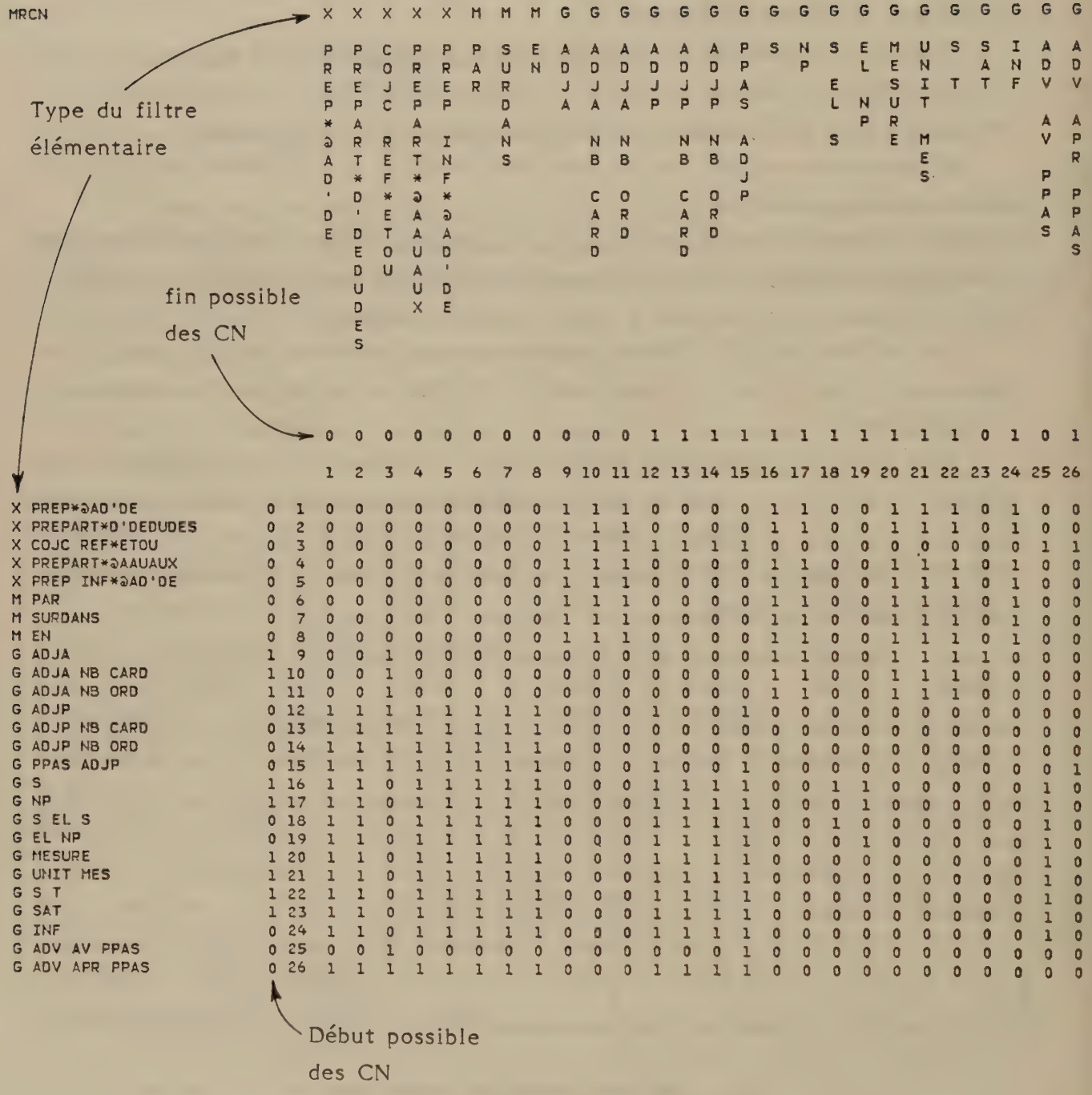
Bien que les expériences décrites dans le présent exposé aient porté sur environ 15 000 mots, nous ne donnons dans le tableau de la figure 15 que les taux d'erreurs et de silence comptabilisés sur une longueur de texte d'environ 1750 mots représentant 80 phrases, car, pour le moment, ces calculs de performances se font à la main.

Nbre de mots	Nbre de phrases	Nbre de CN réelles	Nbre de CN reconnues	Nbre d' erreurs / taux d'erreurs	Silence / taux de silence
1754	80	263	270	7 / 2,6 %	12 / 4,6 %

fig. 15 Performances de l'algorithme de reconnaissance des chaînes nominales.

(*) De même que précédemment, il existe un réseau de transition équivalent, mais cette fois-ci, il est suffisamment complexe pour que l'on se dispense de le reproduire.

'MATRICE DE RECONNAISSANCE DES CHAINES NOMINALES'



MRCN			X PREP*∂AD'DE	X PREPART*D'DEDUDES	X COJC REF*ETOU	X PREPART*∂AAUAUX	X PREP INF*∂AD'DE	M PAR	M SURDANS	M EN	G ADJA	G ADJA NB CARD	G ADJA NB ORD	G ADJP	G ADJP NB CARD	G ADJP NB ORD	G PPAS ADJP	G S	G NP	G S EL S	G EL NP	G MESURE	G UNIT MES	G S T	G SAT	G INF	G ADV AV PPAS	G ADV APR PPAS
			0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
X PREP*∂AD'DE	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	0
X PREPART*D'DEDUDES	0	2	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	0
X COJC REF*ETOU	0	3	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1
X PREPART*∂AAUAUX	0	4	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	0
X PREP INF*∂AD'DE	0	5	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	0
M PAR	0	6	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	0
M SURDANS	0	7	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	0
M EN	0	8	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	0
G ADJA	1	9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	0	0	0
G ADJA NB CARD	1	10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	0	0
G ADJA NB ORD	1	11	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	0	0
G ADJP	0	12	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
G ADJP NB CARD	0	13	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G ADJP NB ORD	0	14	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G PPAS ADJP	0	15	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1
G S	1	16	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0
G NP	1	17	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0
G S EL S	0	18	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0
G EL NP	0	19	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0
G MESURE	1	20	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
G UNIT MES	1	21	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
G S T	1	22	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
G SAT	1	23	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
G INF	0	24	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
G ADV AV PPAS	0	25	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
G ADV APR PPAS	0	26	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0

fig. 12.

```
        PHRASE NO :  1   ELLE COMMENCE AU MOT NO:      1

        % LES VARIATIONS DU FACTEUR DE CONCENTRATION DU COBALT-60 CHEZ LA MOULE &MYTILUS &EDULIS &L. .

.....

******** CHAINES NOMINALES RECONNUES

        LES VARIATIONS DU FACTEUR DE CONCENTRATION DU COBALT-60
        CHEZ LA MOULE &MYTILUS &EDULIS &L.

.....

        PHRASE NO :  2   ELLE COMMENCE AU MOT NO:     17

        $ LE COBALT-60 ENTRE DANS LA COMPOSITION DES EFFLUENTS RADIOACTIFS LI#ES ∂A L' UTILISATION DE L'
        #ENERGIE NUCL#EAIRE .

.....

******** CHAINES NOMINALES RECONNUES

        LE COBALT-60
        DANS LA COMPOSITION DES EFFLUENTS RADIOACTIFS LI#ES ∂A L' UTILISATION DE L' #ENERGIE NUCL#EAIRE

.....

        PHRASE NO : 12   ELLE COMMENCE AU MOT NO:    234

        DANS L' #ETUDE DE TOUTE CONTAMINATION IL CONVIENT DE TENIR COMPTE DE L' IMPORTANCE RELATIVE DE
        PROCESSUS PHYSICO-CHIMIQUES ET BIOLOGIQUES POUVANT INTERVENIR SIMULTAN#EMENT .

.....

******** CHAINES NOMINALES RECONNUES

        DANS L' #ETUDE DE TOUTE CONTAMINATION
        DE L' IMPORTANCE RELATIVE DE PROCESSUS PHYSICO-CHIMIQUES ET BIOLOGIQUES

.....

        PHRASE NO : 26   ELLE COMMENCE AU MOT NO:    681

        POUR R#EDUIRE LES TEMPS DE POSE ET AUGMENTER LA LUMINANCE DE SORTIE ON A CHERCH#E ∂A CONVERTIR LES
        NEUTRONS EN #ELECTRONS ET ∂A AMPLIFIER LE FLUX DES #ELECTRONS DE CONVERSION .

.....

******** CHAINES NOMINALES RECONNUES

        LES TEMPS DE POSE
        LA LUMINANCE DE SORTIE
        LES NEUTRONS EN #ELECTRONS
        LE FLUX DES #ELECTRONS DE CONVERSION

.....

        PHRASE NO : 27   ELLE COMMENCE AU MOT NO:    712

        POUR CECI L' ID#EE A #ET#E D' ASSOCIER UNE GALETTE DE MICROCANAUX ∂A UN CONVERTISSEUR AU GADOLINIUM
        FONCTIONNANT EN TRANSMISSION .

.....

******** CHAINES NOMINALES RECONNUES

        L' ID#EE
        UNE GALETTE DE MICROCANAUX ∂A UN CONVERTISSEUR AU GADOLINIUM
        EN TRANSMISSION

.....
```

fig. 13. Quelques exemples de reconnaissance de CN d'un texte scientifique.

PHRASE NO : 54 ELLE COMMENCE AU MOT NO: 1537

UN NOUVEL ACTE DE LOCATION EST PASS#E PAR L' ORGANISME PROPRI#ETAIRE AVEC CHACUN DES INT#ERESS#ES POUR LES LOCAUX DANS LESQUELS ILS SONT INSTALL#ES .

.....

******* CHAINES NOMINALES RECONNUES

UN NOUVEL ACTE DE LOCATION
PAR L' ORGANISME PROPRI#ETAIRE
AVEC CHACUN DES INT#ERESS#ES
POUR LES LOCAUX

.....

PHRASE NO : 57 ELLE COMMENCE AU MOT NO: 1602

$00000100 LE CONSEIL SUP#ERIEUR DU MINISTƏERE DE LA CONSTRUCTION SIƏEGE AU MINISTƏERE DE LA CONSTRUCTION .

.....

******* CHAINES NOMINALES RECONNUES

LE CONSEIL SUP#ERIEUR DU MINISTƏERE DE LA CONSTRUCTION
AU MINISTƏERE DE LA CONSTRUCTION

.....

PHRASE NO : 62 ELLE COMMENCE AU MOT NO: 1681

LE CONSEIL SUP#ERIEUR #ELIT SON VICE-PR#ESIDENT PARMI SES MEMBRES .

.....

******* CHAINES NOMINALES RECONNUES

LE CONSEIL SUP#ERIEUR
SON VICE-PR#ESIDENT
PARMI SES MEMBRES

.....

PHRASE NO :107 ELLE COMMENCE AU MOT NO: 3590

CETTE MAJORATION EST AUGMENT#EE D' UN POINT POUR CHAQUE ANN#EE SUPPL#EMENTAIRE DE S#EPARATION .

.....

******* CHAINES NOMINALES RECONNUES

CETTE MAJORATION
D' UN POINT
POUR CHAQUE ANN#EE SUPPL#EMENTAIRE DE S#EPARATION

.....

PHRASE NO :139 ELLE COMMENCE AU MOT NO: 4655

EN CAS DE N#ECESSIT#E ABSOLUE DE CR#EATION D' UN D#EGAGEMENT SUR LE VESTIBULE , IL CONVIENT D' INTERPOSER UN SAS ƏA DEUX PORTES PLEINES ƏA FERMETURE AUTOMATIQUE .

.....

******* CHAINES NOMINALES RECONNUES

EN CAS DE N#ECESSIT#E ABSOLUE DE CR#EATION D' UN D#EGAGEMENT SUR LE VESTIBULE
UN SAS ƏA DEUX PORTES PLEINES ƏA FERMETURE AUTOMATIQUE

.....

fig. 14. Quelques exemples de reconnaissance de CN d'un texte juridique.

On peut encore noter, à titre indicatif, qu'une annulation du taux de silence (consécutive à une acceptation de toutes les prépositions) se traduirait par un accroissement relativement important du taux d'erreurs, lequel passerait en effet de ~2,6 % à ~5 %.

Ce dernier aspect constitue une justification à postériori de l'approximation que nous avons faite quant au choix des prépositions.

III. 2. Reconnaissance des chaînes verbales•:

Tout comme les chaînes nominales, les chaînes verbales peuvent être caractérisées d'une part, par la nature des éléments qui les composent, d'autre part, par la syntaxe de ces éléments.

Elles sont, en effet, dans leur quasi-totalité construites à partir des seules catégories suivantes :

- les verbes à l'infinitif, conjugués, au participe passé ou au participe présent (c'est à dire toutes les formes dérivées) ;
- les adverbes ;
- les adjectifs attributs ;
- les particules préverbales ;
- les prépositions ;
- les conjonctions.

Tout agencement de ces catégories, quand bien même issu d'un texte, ne constitue cependant pas toujours une chaîne verbale.

Ainsi, dans la phrase suivante :

Les billets que vous possédez portent un numéro gagnant

le segment "possédez portent" représente deux chaînes verbales juxtaposées et non pas une seule, malgré la conformité de ses éléments.

Nous nous trouvons donc ainsi en présence de problèmes analogues en tous points à ceux rencontrés lors de la reconnaissance des chaînes nominales.

Seule, en effet, la nature des filtres élémentaires à mettre en oeuvre change.

Aussi donnerons nous d'emblée la matrice qui simulera le filtre P_V de reconnaissance des chaînes verbales (cf. fig. 16) accompagnée de quelques exemples de reconnaissance effectuée sur les mêmes textes que précédemment (cf. fig. 17).

Les remarques et commentaires qui ont été faits restent évidemment valables. Nous laissons au lecteur le soin d'effectuer les transpositions qui s'imposent à leur égard.

Le tableau de la figure 18 donne les performances de l'algorithme de reconnaissance des chaînes verbales.

Les taux de non-reconnaissance ont été comptabilisés sur d'abord 80 phrases (qui ont servi aux calculs précédents ; cf. fig. 15) puis sur 52 autres phrases correspondant à environ 1770 mots.

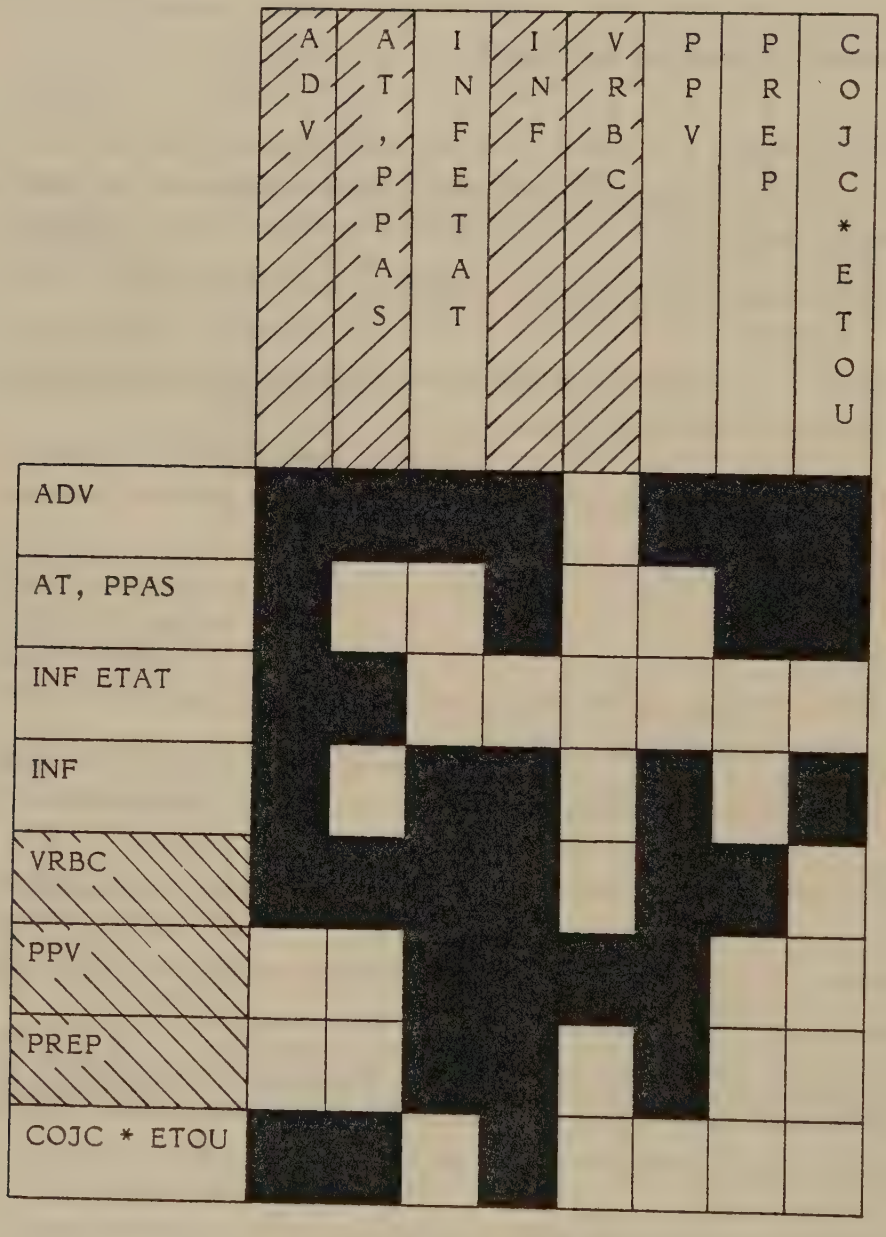


fig. 16. Matrice de reconnaissance des chaînes verbales. (voir note bas de page pour les abréviations).

ADV : ADVerbe ; AT : ATtribut ; PPAS : Participe PASsé ; INF ETAT : INFinitif ETAT ; INF : INFinitif ; VRBC : VeRBe Conjugué ; PPV : Particule PréVerbale ; PREP : PREPosition ; COJC : COnJonction de Coordination.

Les figures 17.a et 17.b donnent quelques exemples de phrases ayant été soumises aux filtres de reconnaissance des chaînes nominales et verbales.

La phrase numéro 66 offre notamment un exemple où la reconnaissance de la chaîne verbale n'est pas achevée.

En effet, l'algorithme reconnaît deux chaînes verbales : "peut" et "décider", qui, en fait, n'en constituent qu'une seule : "peut décider".

	Nbre de mots	Nbre de phrases	Nbre de CV réelles	Nbre de CV reconnues	Nbre de non-rec. / Taux de non-reconnaissance
	1754	80	111	112	1
	1771	52	103	105	2
Total :	3525	132	214	217	3 / ~ 1,4 %

fig. 18. Performances de l'algorithme de reconnaissance des chaînes verbales.

Ainsi que nous pouvons le voir, cette non-reconnaissance (comptabilisée dans le premier lot des 80 phrases ; cf. fig. 18) est due à l'insertion dans la chaîne verbale d'une locution adverbiale placée entre virgules.

La reconnaissance de la chaîne verbale nécessite donc ici une pré-étape (*) durant laquelle aurait été reconnue puis élidée la locution adverbiale "en outre" ("ajout" dans les grammaires en chaînes, SALKOFF, 103).

Cette solution doit en fait être généralisée.

(*) ou encore un post-traitement qui aurait pour objet de "ressouder" deux chaînes verbales dans certaines conditions.

```
PHRASE NO :  4   ELLE COMMENCE AU MOT NO:    82

CE RADIONUCL#EIDE EST SUSCEPTIBLE DE SE FIXER SUR LES PARTICULES EN SUSPENSION QUI SONT PI#EG#EES
PAR LES ESPDECES FILTREUSES TELLES QUE LES MOULES ( &MYTILUS &EDULIS &L. ) .

.....

******* CHAINES NOMINALES RECONNUES

CE RADIONUCL#EIDE
SUR LES PARTICULES EN SUSPENSION
PAR LES ESPDECES FILTREUSES
LES MOULES

.....

******* CHAINES VERBALES RECONNUES

EST SUSCEPTIBLE DE SE FIXER                                    S FIXER
SONT PI#EG#EES                                                 P PI#EG#EES


PHRASE NO : 12   ELLE COMMENCE AU MOT NO:   234

DANS L' #ETUDE DE TOUTE CONTAMINATION IL CONVIENT DE TENIR COMPTE DE L' IMPORTANCE RELATIVE DE
PROCESSUS PHYSICO-CHIMIQUES ET BIOLOGIQUES POUVANT INTERVENIR SIMULTAN#EMENT .

.....

******* CHAINES NOMINALES RECONNUES

DANS L' #ETUDE DE TOUTE CONTAMINATION
DE L' IMPORTANCE RELATIVE DE PROCESSUS PHYSICO-CHIMIQUES ET BIOLOGIQUES

.....

******* CHAINES VERBALES RECONNUES

CONVIENT DE TENIR COMPTE                                       S TENIR COMPTE
POUVANT INTERVENIR SIMULTAN#EMENT                              S INTERVENIR


PHRASE NO : 52   ELLE COMMENCE AU MOT NO:  1497

L' #ECHANGE NE PEUT -ETRE IMPOS#E AUX LOCATAIRES QUI ONT #ET#E SINISTR#E DANS LES LIEUX LOU#ES .

.....

******* CHAINES NOMINALES RECONNUES

L' #ECHANGE
AUX LOCATAIRES
DANS LES LIEUX LOU#ES

.....

******* CHAINES VERBALES RECONNUES

NE PEUT -ETRE IMPOS#E                                          P IMPOS#E
ONT #ET#E SINISTR#E                                            P SINISTR#E
```

fig. 17.(a). Exemples de segmentation de phrases : reconnaissance des chaînes nominales et verbales.

PHRASE NO : 56 ELLE COMMENCE AU MOT NO: 1577

TOUTEFOIS , CES DISPOSITIONS NE SERONT PAS OPPOSABLES AUX PERSONNES QUI AURONT D#EPOS#E UNE DEMANDE DE PERMIS DE CONSTRUIRE AVANT LE 1ER JUIN 1959 .

.....

******* CHAINES NOMINALES RECONNUES

CES DISPOSITIONS
AUX PERSONNES
UNE DEMANDE DE PERMIS DE CONSTRUIRE

.....

******* CHAINES VERBALES RECONNUES

NE SERONT PAS OPPOSABLES	P OPPOSABLES
AURONT D#EPOS#E	S D#EPOS#E

PHRASE NO : 66 ELLE COMMENCE AU MOT NO: 1751

LE CONSEIL SUP#ERIEUR PEUT , EN OUTRE , D#ECIDER LA CR#EATION D' AUTRES SECTIONS DE CARACTƏERE TEMPORAIRE POUR L' EXAMEN DE PROBLƏEMES PARTICULIERS .

.....

******* CHAINES NOMINALES RECONNUES

LE CONSEIL SUP#ERIEUR
LA CR#EATION D' AUTRES SECTIONS DE CARACTƏERE TEMPORAIRE
POUR L' EXAMEN DE PROBLƏEMES PARTICULIERS

.....

******* CHAINES VERBALES RECONNUES

PEUT	S PEUT
D#ECIDER	S D#ECIDER

PHRASE NO : 73 ELLE COMMENCE AU MOT NO: 1948

CE LOGEMENT EST CONSID#ER#E COMME #EQUIP#E EN EAU , GAZ , #ELECTRICIT#E ; L' IMMEUBLE N' EST PAS DOT#E D' ASCENSEUR .

.....

******* CHAINES NOMINALES RECONNUES

CE LOGEMENT
EN EAU
GAZ
#ELECTRICIT#E
L' IMMEUBLE
D' ASCENSEUR

.....

******* CHAINES VERBALES RECONNUES

EST CONSID#ER#E COMME #EQUIP#E	P #EQUIP#E
N' EST PAS DOT#E	P DOT#E

PHRASE NO :108 ELLE COMMENCE AU MOT NO: 3604

$00000810 SI UN FOYER HABITE UN LOGEMENT TROP #ELOIGN#E DU LIEU DE TRAVAIL DU CHEF DE FAMILLE , LA NOTE AFF#ERENTE ƏA CE FOYER EST MAJOR#EE DE 2 POINTS .

.....

******* CHAINES NOMINALES RECONNUES

UN FOYER
UN LOGEMENT TROP #ELOIGN#E DU LIEU DE TRAVAIL DU CHEF DE FAMILLE
LA NOTE AFF#ERENTE ƏA CE FOYER
DE 2 POINTS

.....

******* CHAINES VERBALES RECONNUES

HABITE	S HABITE
EST MAJOR#EE	P MAJOR#EE

fig. 17.(b).

En effet, les deux autres cas de non-reconnaissance (cf. fig. 18) sont dus à l'insertion dans la chaîne verbale, non pas d'une locution adverbiale comme plus haut, mais d'un syntagme prépositionnel placé entre deux virgules. La figure 19 en donne un exemple.

```
PHRASE NO :106   ELLE COMMENCE AU MOT NO:  3545

$00000800 SI LES DIFF#ERENTS MEMBRES D' UN M~EME FOYER DEMANDEUR SONT CONTRAINTS , PAR L'
INSUFFISANCE DE LOGEMENT , DE VIVRE DEPUIS PLUS D' UN AN DANS DES COMMUNES DIFF#ERENTES , LA NOTE
AFF#ERENTE AU FOYER DU CHEF DE FAMILLE EST MAJOR#EE DE 4 POINTS .

.....

******* CHAINES NOMINALES RECONNUES

        LES DIFF#ERENTS MEMBRES D' UN M~EME FOYER DEMANDEUR
        PAR L' INSUFFISANCE DE LOGEMENT
        DEPUIS PLUS D' UN AN DANS DES COMMUNES DIFF#ERENTES
        LA NOTE AFF#ERENTE AU FOYER DU CHEF DE FAMILLE
        DE 4 POINTS

.....

******* CHAINES VERBALES RECONNUES

        SONT CONTRAINTS                          P CONTRAINTS
        DE VIVRE                                 S VIVRE
        EST MAJOR#EE                             P MAJOR#EE
```

fig. 19. Un exemple de non reconnaissance de chaîne verbale.

Il y a donc lieu, dans cette pré-étape, de reconnaître puis d'élider aussi bien les locutions adverbiales que les syntagmes prépositionnels situés entre virgules.

Notons que les algorithmes de reconnaissance restent les mêmes que ceux qui ont été développés pour la reconnaissance des chaînes nominales ou verbales : seules changent les matrices de reconnaissance.

Cette solution ne pourrait évidemment convenir dans le cas d'insertion, dans les chaînes verbales, de structures complexes incluant notamment des verbes. En effet, il conviendrait alors d'introduire une récursivité dans l'utilisation des diverses matrices de reconnaissance.

Remarquons néanmoins la rareté de ces cas dans les textes traités.

La segmentation de la phrase en chaînes nominales et verbales étant ainsi atteinte, nous sommes en mesure, désormais, de détailler quelque peu la "brique" *Traitements syntaxico-sémantiques* de la figure 7 (cf. fig. 20).

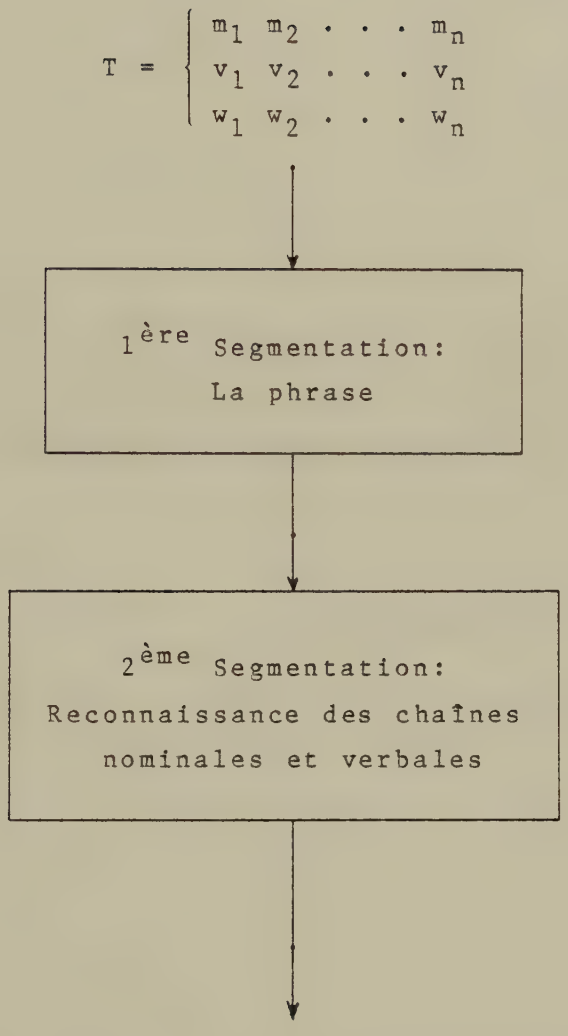


fig. 20. Segmentation de la phrase en chaînes nominales et verbales

deuxième partie

CONSTRUCTION

DES RELATIONS DE DEPENDANCE POTENTIELLES

IV. Analyse sous forme de dépendances :

Notre but est d'obtenir d'emblée les divers arcs de dépendance qui s'établissent entre les différents éléments d'une même phrase.

Ainsi, à la phrase :

" un nouvel acte de location est passé par l'organisme propriétaire avec les intéressés"

(Règlementation de l'urbanisme).

on veut associer la représentation :

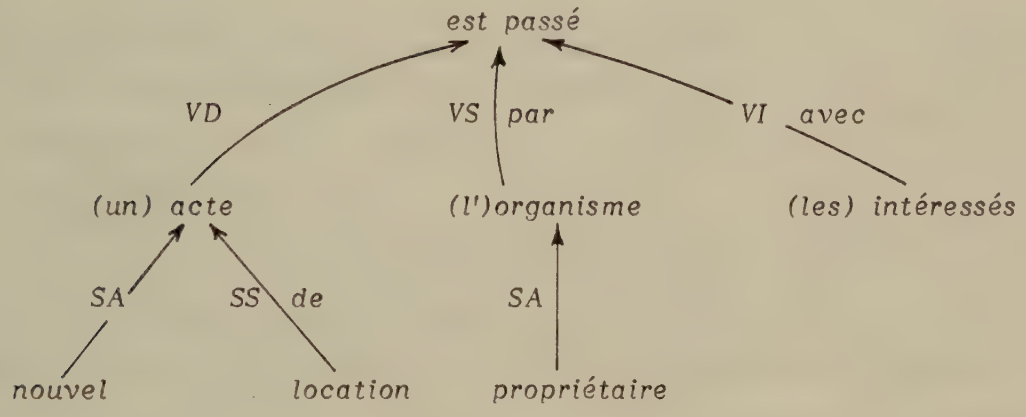


fig. 21.

Or, comme nous l'avons vu, les phrases soumises à l'analyse conduisent souvent à des représentations parasites que seule une connaissance sémantique peut éléminer (cf. ch. I).

Notre but sera donc, pour l'instant, d'obtenir toutes les représentations potentielles susceptibles d'être associées aux phrases soumises à l'analyse, laissant ainsi aux étapes ultérieures le soin d'éliminer les représentations parasites éventuelles.

Ainsi, si nous reprenons la phrase :

Vente des produits de beauté aux algues marines

nous voulons avoir, à l'issue de la présente étape, les trois représentations possibles suivantes :

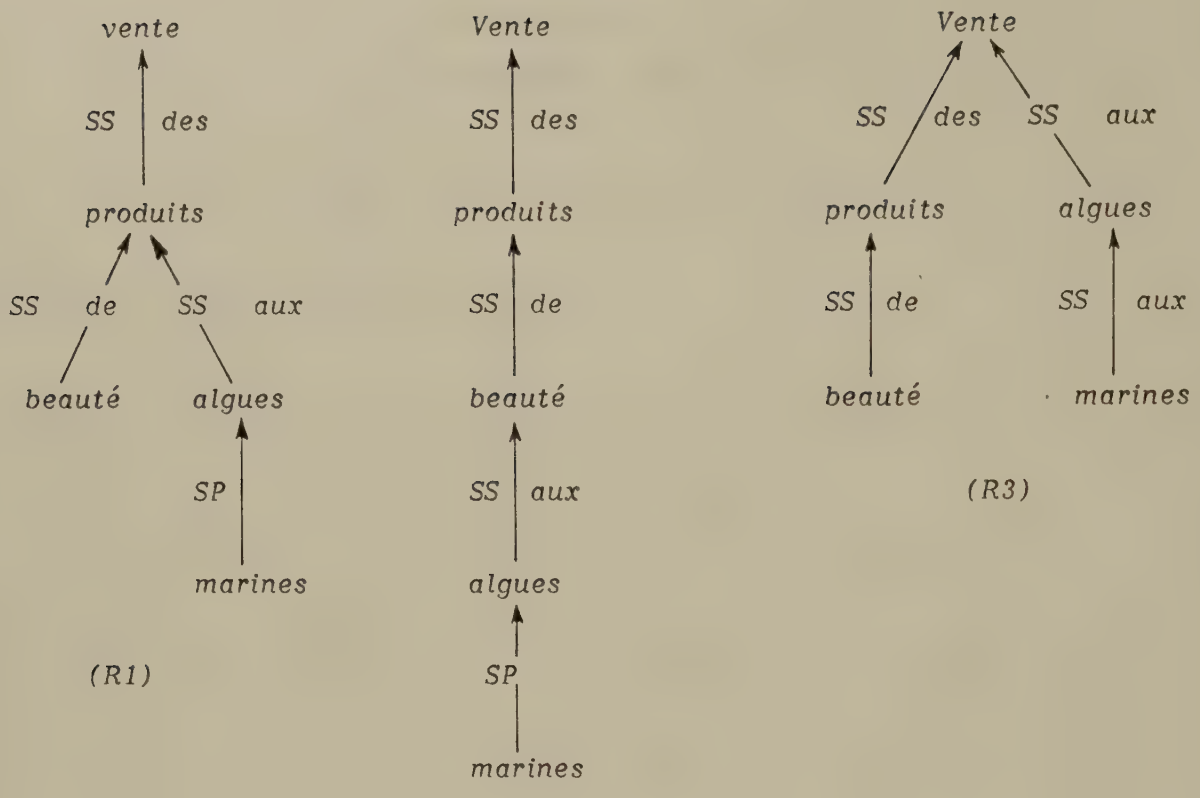


fig. 22

dont une seule est pertinente, à savoir la représentation (R1).

Remarquons cependant dès maintenant que l'aspect séquentiel des traitements (syntaxiques puis sémantiques) n'est en fait qu'apparent. Nous verrons en effet qu'il y a imbrication dans les traitements et que la composante sémantique intervient parallèlement lors de la construction des diverses relations de dépendance (cf. ch. V, § IV. 3).

IV. 0. Bruit - Silence•:

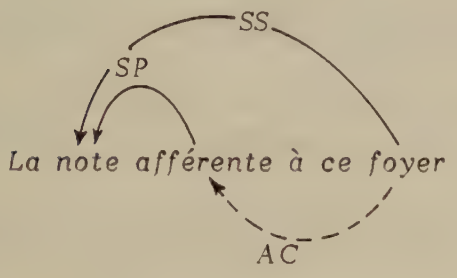
Afin de rendre commode la description des résultats obtenus nous introduirons deux vocables visant à mesurer les performances des algorithmes proposés et expérimentés.

Le but étant de construire, à cette étape, l'ensemble des relations de dépendance potentielles, nous dirons que l'on est en présence :

• d'un bruit si la relation potentielle établie est en fait appelée à disparaître lors du traitement sémantique ultérieur ;

• d'un silence si la relation convoitée ne se trouve pas en fait dans le lot des relations potentielles établies.

Ainsi dans la phrase suivante :



fléchée comme indiquée, nous dirons que l'on est en présence :

- d'un bruit relativement aux relations SS_p : la relation (note, $SS_à$, foyer) est en effet erronée et est appelée à disparaître lors du traitement sémantique ;
- d'un silence relativement aux relations AC_p : la relation (afférente, $AC_à$, foyer) est en effet recherchée mais non établie.

Dès lors, quatre situations différentes peuvent se présenter selon que l'on a ou non du bruit ou du silence. Ces situations peuvent être résumées dans le tableau suivant où 1 (respectivement 0) indique la présence (respectivement l'absence) du bruit ou du silence.

Situation n°	Silence	Bruit	Commentaires
1	0	0	Les relations potentielles obtenues sont d'emblée correctes. C'est la situation idéale. il n'y a pas d'ambiguïté.
2	0	1	Les relations convoitées se trouvent parmi les les relations potentielles établies. C'est la situation ambiguë.
3	1	0	Aucune relation potentielle n'est établie
4	1	1	Les seules relations potentielles établies sont erronées.

fig. 23.

Nous verrons que ce tableau est à la base de l'apprentissage sémantique.

Nous verrons également que si la situation 3 n'est pas source d'erreurs dans cet apprentissage, la situation 4 le sera par contre, et à cet égard devra être évitée chaque fois que cela sera possible.

Nous dirons de ces deux dernières situations que la 3 procure un silence total et la 4 un bruit total ou encore un silence partiel.

Les options que nous serons amenés à prendre devront donc conduire aux situations 1 ou 2, mais en tout état de cause éviter les situations 3 et, surtout, 4.

Nous serons amenés en fait à faire un compromis et à nous accommoder quelquefois des situations 3 et 4.

Nous avons cependant fixé un seuil au delà duquel toute option serait rejetée si elle devait avoir pour incidence un taux de bruit total supérieur à ce seuil.

Celui-ci a été fixé arbitrairement à 10 %. En particulier, nous préférerons une absence d'analyse (silence total) à une analyse dont le bruit total, relativement à une relation donnée, serait supérieur à ce seuil.

Notons enfin, que ce sera généralement la conjonction de deux ou plusieurs options qui conduiront aux diverses situations ci-dessus énumérées.

IV. 1. Relations satellites :

Selon la nature des mots (pleins ou vides[*]) nous distinguerons deux types de relations de dépendance :

- d'une part, celles qui s'établissent entre mots vides et mots pleins,
- d'autre part, celles qui s'établissent entre mots pleins.

Suivant en cela L. TESNIERE, nous utiliserons le terme "satellite", adjectivé, pour désigner les premières.

Faisant correspondre à la distinction entre les mots pleins et les mots vides (sur le plan sémantique) celle entre les mots constitutifs et les mots subsidiaires (sur le plan structural) (108, p. 55) , L. TESNIERE dit en effet: *"Les mots subsidiaires sont donc en quelque sorte comme des satellites qui seraient tombés sur l'astre (mots constitutifs) autour duquel ils gravitaient".* (108, p. 57) .

(*) Retenons à leur sujet qu'ils"sont en définitive, les éléments les plus évolués, donc les plus humanisés et les plus intéressants du langage" (L.TESNIERE 108, p. 81)

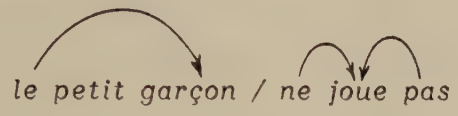
Aussi dirons-nous "relations satellites" pour les premières, les secondes correspondant aux RLS syntagmatiques que nous nous proposons d'obtenir.

Notons que la construction des premières constitue une pré-étape nécessaire pour la construction des secondes.

La distinction homo/hétéro-syntagmatique peut là aussi être introduite.

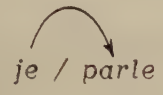
Il est possible, en effet, de remarquer que, outre les relations satellites qui s'établissent entre éléments appartenant à une même chaîne (nominale ou verbale),

Exemple :



d'autres s'établissent entre éléments appartenant à des chaînes différentes.

C'est par exemple le cas des relations qui s'établissent entre le sujet pronominal et le verbe. Par exemple :



C'est aussi souvent le cas des relations qui s'établissent entre les pronoms personnels compléments et le verbe (L. TESNIERE, 108) .

Cependant, qu'il s'agisse de relations homosyntagmatiques ou de relations hétérosyntagmatiques, leur détermination reste dans la plupart des cas simple : les mots à relier sont en effet presque toujours contigus,(si ce n'est, quelquefois, quasiment agglutinés ; ex : "donnez-le lui" (108))

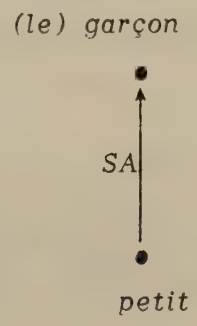
Une autre distinction peut être introduite entre les relations satellites, fondée cette fois sur la nature des mots vides mis en jeu.

En effet certains d'entre eux ont une fonction anaphorique. *"ils deviennent pleins dès que, dans une phrase, ils entrent en connexion anaphorique avec un autre mot, parce que celui-ci leur communique son sens... Ce sont des mots pleins virtuels"* (108, p. 86 & 90) .

Il leur correspondra donc, dans le graphe, un noeud qui sera appelé par la suite à recevoir la valeur de la "source sémantique" qui leur est associée (cf. ch. III, § IV. 2. 4.). Par contre les autres mots vides seront simplement rattachés directement aux noeuds correspondant aux mots pleins auxquels ils se rapportent.

Une dernière dichotomie sera enfin effectuée entre les relations satellites mettant en jeu des mots vides n'ayant pas de fonction anaphorique. Graphiquement ils seront en effet :

- ou mis entre parenthèses et ils précèderont alors les mots auxquels ils se rapportent c'est le cas notamment des articles ou de la négation ; par exemple "le petit garçon" recevra la représentration suivante :



- ou bien encore placés à proximité des flèches représentant les relations LS syntagmatiques, ils servent alors à étiqueter les relations de dépendance qui s'établissent entre mots pleins. C'est le cas des prépositions par exemple.(cf. ch. II. § III. 2.).

Cette typologie des diverses relations satellites peut être résumée par la figure 24 :

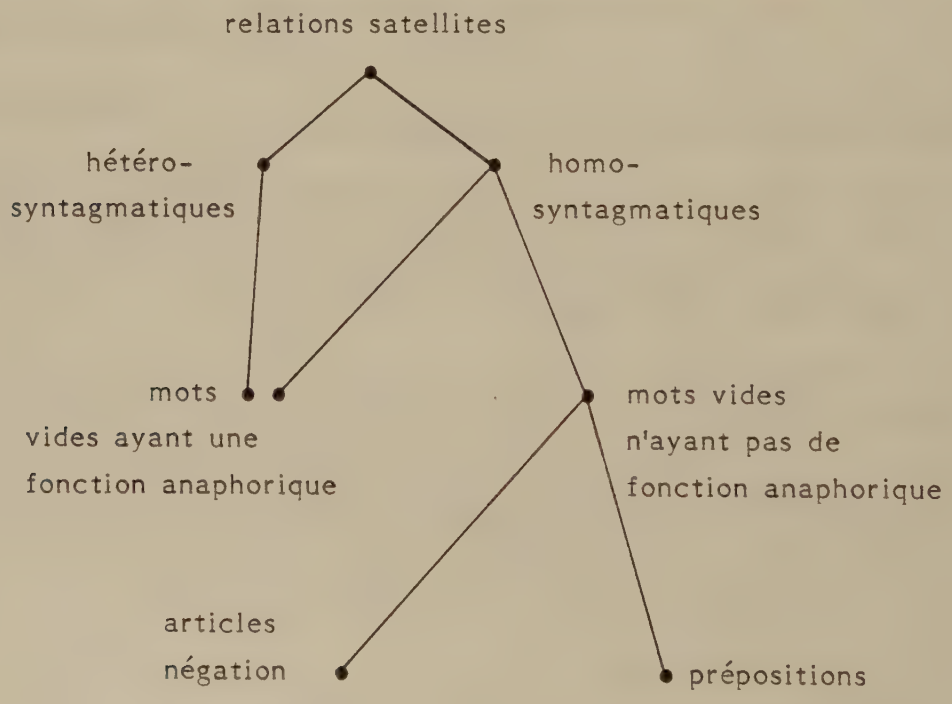


fig. 24 Typologie des diverses relations satellites

IV. 2. Reconnaissance des relations de dépendance :

Chaque relation de dépendance peut être caractérisée par :

- le type de liaison auquel elle appartient : c'est par exemple l'une des différentes RLS syntagmatiques définies dans le chapitre II ou encore l'une des relations satellites vues plus haut ;
- les conditions sur chacun des éléments qui la constituent. Ces conditions peuvent être morphologiques (genre-nombre), grammaticales, syntaxiques (elles tiennent compte aussi bien de l'ordre relatif, dans le texte, des éléments à relier que de leur contexte) ou, enfin, sémantiques (compatibilité, cf. ch. V).

Rappelons que notre but est d'obtenir, à ce stade, toutes les relations de dépendance potentielles sans qu'intervienne encore la composante sémantique. Leur reconnaissance sera donc, pour le moment, fondée sur l'unique utilisation de filtres syntaxiques.

Les filtres élémentaires permettront en effet la reconnaissance des éléments constitutifs des diverses relations de dépendance (cette reconnaissance inclut notamment la vérification de certaines conditions morpho-grammaticales) ou de leur environnement ; leur agencement permettant de prendre en compte les conditions syntaxiques relatives aux éléments mis en relation ou à leur contexte.

La forme générale de ces filtres sera donc la suivante :

$$P_x = P_1\ F_A\ P_2\ F_B\ P_3$$

où :

- P_x est le filtre de reconnaissance de la relation de type x que l'on souhaite obtenir ; c'est le filtre à définir ;
- P_1 , P_2 et P_3 sont des filtres syntaxiques permettant de prendre en compte les conditions contextuelles relatives aux éléments de la relation. P_2 permet notamment d'instaurer les barrières ou frontières spécifiques de la relation x à ne pas franchir (cf. ch. III, § II. 3.) ;
- F_A et F_B permettent enfin la reconnaissance des éléments constitutifs de la relation.

IV. 2. 1. Analyse de la chaîne nominale :

Les chaînes nominales ayant été reconnues lors de l'étape précédente, nous allons procéder maintenant à leur analyse syntaxique, c'est à dire à la détermi-

nation de toutes les relations de dépendance potentielles qui s'y trouvent.

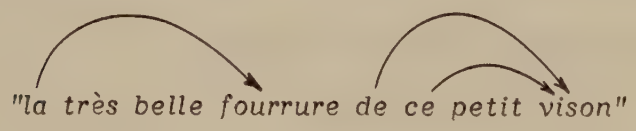
IV. 2. 1. 1. Construction des relations satellites :

Ainsi que nous l'avons vu, la reconnaissance des relations satellites est relativement simple : les liaisons à établir ne donnent pas lieu à des ambiguïtés, les éléments à relier étant presque toujours contigus. Par ailleurs, la reconnaissance de ces éléments est aussi relativement aisée : il s'agit en effet de reconnaître d'une part les mots pleins, en l'occurrence les substantifs ; d'autre part les mots vides, en l'occurrence les articles et/ou prépositions.

La définition des filtres syntaxiques permettant la reconnaissance de ces relations satellites est dès lors simple. L'allure générale de ces filtres sera la suivante :

$$P_{SAT} = F_V \quad P^{cont}_{SAT} F_P$$

où : F_V est le filtre élémentaire de reconnaissance du mot vide,
F_P le filtre élémentaire de reconnaissance du mot plein,
et F^{cont}_{SAT} un filtre syntaxique permettant de prendre en compte les cas de non contiguïté ; par exemple :



Les éléments susceptibles de s'intercaler entre les éléments à relier ne peuvent en effet prendre leur valeur que dans un ensemble déterminé, dépendant de la nature de la relation. On ne peut trouver par exemple entre un "article" et un "substantif" une "préposition".

Les configurations qu'il est permis de rencontrer entre les éléments à relier peuvent donc être reconnues au moyen d'un filtre syntaxique analogue, dans sa conception, aux filtres de reconnaissance des chaînes nominales ou verbales.

La simplicité apparente de la reconnaissance des relations satellites ne doit cependant pas éclipser l'importance qu'elles revêtent pour les traitements à venir.

En effet, comme nous le verrons, les filtres syntaxiques ultérieurs de reconnaissance des RLS syntagmatiques devront tenir compte du contexte des mots à relier : il s'agit dans le cas présent de spécifier la nature du contexte immédiat gauche.

La reconnaissance des relations satellites s'accompagnera donc d'une opération d'ajout au compte du mot plein, d'une information relative à ce contexte gauche.

C'est pourquoi nous introduirons dès maintenant une double distinction entre les substantifs : selon qu'ils sont déterminés (présence d'un article) ou non, d'une part ; et selon la nature de la préposition qui leur est éventuellement rattachée, d'autre part.

Nous donnons ci-après, à titre indicatif, la définition des filtres élémentaires qui ont été retenus pour la détermination des diverses relations satellites internes aux chaînes nominales. Leur justification complète ne pourra cependant être fournie que d'une manière répartie dans les paragraphes suivants.

a) filtres de reconnaissance des mots vides.

Bien que pour cette étape, nous puissions nous contenter de définir deux types de filtres élémentaires (articles et prépositions), la nécessité d'établir ultérieurement diverses distinctions, effleurées ci-dessus, nous amène à définir les filtres élémentaires de reconnaissance des mots vides comme suit :

1°) Les articles :

- sans fonction anaphorique :

$$F_V^{Art} = \langle \textit{articles définis, articles indéfinis, adjectifs démonstratifs, adjectifs indéfinis} \rangle$$

- avec fonction anaphorique :

$$F_V^{Artpos} = \langle \textit{adjectifs possessifs, adjectifs relatifs} \rangle$$

2°) Les prépositions :

$$F_V^{Prep1} = \langle (d', PREPART, *), (de, PREPART, *), (à, PREPART, *), (en, PREPART, *) \rangle$$

$$F_V^{Prep2} = \langle (d', PREP, *), (de, PREP, *), (à, PREP, *), (du, PREP, *) (des, PREPART, *), (au, PREPART, *), (aux, PREPART, *) \rangle$$

$F_V^{Prep3} = \langle (par, *, *) \rangle$

Les prépositions ou prépositions-articles n'apparaissant pas dans l'un des trois filtres précédents définiront enfin une quatrième classe. Ils constitueront, comme nous le verrons des "barrières" pour certaines relations LS (cf. ch. III, § II. 3. et § IV. 2. 1. 2.).

b) filtres de reconnaissance des mots pleins :

Seuls les substantifs sont susceptibles de participer à des relations satellites dans les chaînes nominales. Leur reconnaissance se fera donc au moyen du filtre élémentaire suivant, déjà défini (cf. ch. III, § III. 1. 2. 1.)

$F^S = \langle$ Substantifs $\rangle$

Nous retiendrons qu'à l'issue de la présente étape, nous sommes en présence de diverses distinctions quant aux substantifs, que l'on peut représenter par l'arbre suivant:

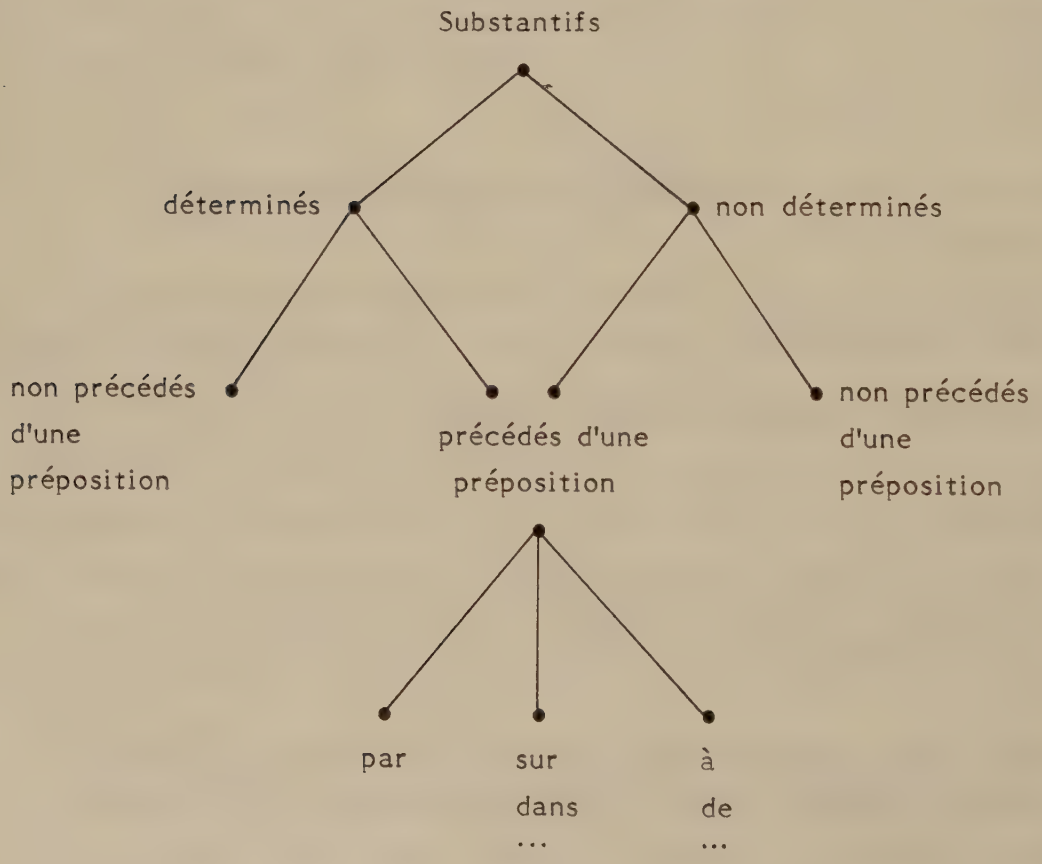


fig. 25.

Cet arbre marque, comme nous le verrons, les différentes potentialités qu'ont les substantifs d'être sujet, compléments de verbes, de noms ou d'adjectifs.

La figure suivante donne des exemples de reconnaissance automatique de relations satellites effectuées sur des chaînes nominales obtenues lors de l'étape précédente.

Nous pouvons notamment y voir que mots vides et mots pleins ont été reconnus et que ces derniers ont, en outre, reçu un numéro d'ordre qui servira dans la suite à indiquer les liaisons ou fléchages.

Les relations satellites sont représentées sur une même ligne. Ainsi dans "en petite quantité" (cf. fig. 26, 1ère chaîne nominale) "en" est rapporté à "quantité" et est donc imprimé sur la même ligne.

Remarquons enfin que ce sont ces relations qui, très souvent, permettent la détermination des valeurs morphologiques du genre et du nombre.

Exemple :

- *Genre : Le physique , la physique*
  Un tour , une tour

- *Nombre : Un propos , des propos*
  Le choix , les choix

IV 2. 1. 2. Construction des RLS homosyntagmatiques potentielles.

Il s'agit maintenant d'établir toutes les RLS potentielles internes à la chaîne nominale.

Comme nous l'avons vu, toutes les RLS définies dans le chapitre III peuvent apparaître dans les chaînes nominales. Aussi allons-nous les reprendre, en donnant pour chacune les règles d'agencement des filtres élémentaires permettant de les obtenir.

a) RLS Substantif-adjectif Antérieur (SA) :

La détermination de cette relation est, à l'image des relations satellites, relativement simple. On constate en effet que les adjectifs antérieurs ont tendance *"à s'agglutiner au substantif subséquent pour former un nouveau substantif"* (TESNIERE, 108, p. 148) . Les cas de non-contiguïté sont rares, et quand ils se produisent ils ne revêtent souvent qu'une forme simple : c'est notamment l'in-

LA FIXATION SUR LES ORGANISMES VIVANTS D' #EL#EMENTS PR#ESENTS EN PETITES QUANTIT#ES DANS L' EAU DE MER

261	1		LA	FIXATION
264	2	SUR	LES	ORGANISMES
265	3			VIVANTS
267	4	D'		#EL#EMENTS
268	5			PR#ESENTS
270	6			PETITES
271	7	EN		QUANTIT#ES
274	8	DANS	L'	EAU
276	9	DE		MER

UNE CONTRIBUTION PASSIVE ƏA LA FIXATION DE L' #EL#EMENT

298	1		UNE	CONTRIBUTION
299	2			PASSIVE
302	3	ƏA	LA	FIXATION
305	4	DE	L'	#EL#EMENT

#ETUDE DE LA FIXATION DU RUTH#ENIUM-106 SUR DIVERS TYPES DE S#EDIMENTS

398	1			#ETUDE
401	2	DE	LA	FIXATION
403	3	DU		RUTH#ENIUM-106
406	4	SUR	DIVERS	TYPES
408	5	DE		S#EDIMENTS

LA PR#ESIDENCE DES SECTIONS R#EUNIES

1809	1		LA	PR#ESIDENCE
1811	2	DES		SECTIONS
1812	3			R#EUNIES

POUR L' #ETUDE DE CERTAINES QUESTIONS PARTICULIƏERES

1852	1	POUR	L'	#ETUDE
1855	2	DE	CERTAINES	QUESTIONS
1856	3			PARTICULIƏERES

AVEC L' AUTORISATION DES MINISTRES

1883	1	AVEC	L'	AUTORISATION
1885	2	DES		MINISTRES

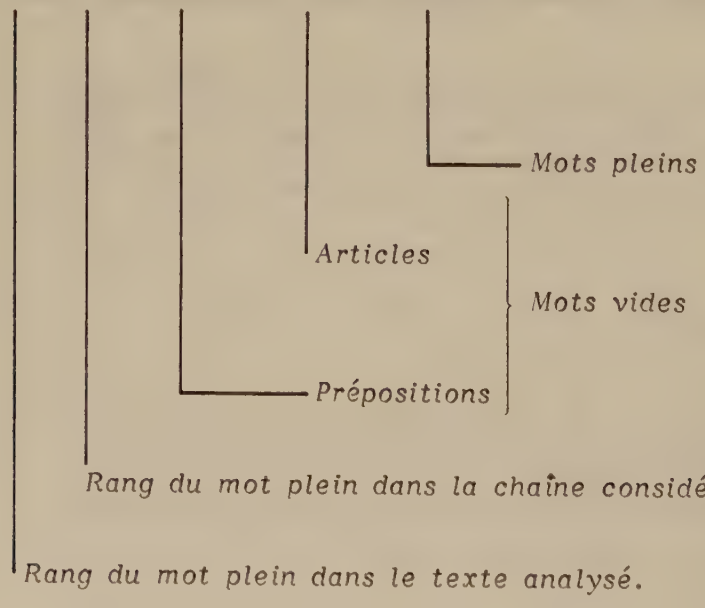
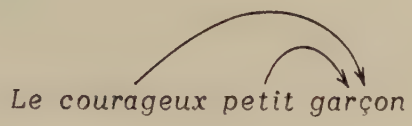


fig. 26 Quelques exemples de reconnaissance des relations satellites (extraits de textes scientifiques et juridiques)

sertion d'autres adjectifs antérieurs.

Exemple :



La reconnaissance des relations SA ne conduit donc pas des ambiguïtés comme cela sera le cas pour les relations SP ou SS_p, par exemple (voir plus loin)

La définition du filtre syntaxique P_{SA} de reconnaissance des relations SA est dès lors la suivante :

$$P_{SA} = F^{AdjA}\ p^{cont}_{SA} F^{S}$$

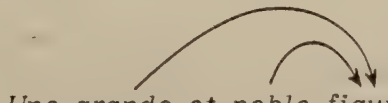
où :

- F^S est le filtre élémentaire de reconnaissance des substantifs ; (cf. ch. III, § III. 1. 2. 1.) ;

- F^{AdjA} le filtre de reconnaissance des adjectifs antérieurs ; (cf. ch. III, § III. 1. 2. 1.) ;

- p^{Cont}_{SA} le filtre syntaxique permettant de prendre en compte les cas de non-contiguïté : on ne peut trouver entre un adjectif antérieur et le substantif auquel il se rapporte que d'autres adjectifs antérieurs, soit juxtaposés, soit coordonnés, éventuellement modifiés par des adverbes.

Exemple :



La figure suivante donne des exemples de reconnaissance de relations SA effectuée sur des textes scientifiques et juridiques (cf. fig. 27).

Remarques :

- Bien que le substantif et l'adjectif qui lui est rapporté aient tendance dans le cas présent à former "un mot unique de sens nouveau" (108, p.149) , il est néanmoins souhaitable de conserver la possibilité de scinder de telles relations. Nous verrons en effet que cela permet d'accroître la capacité inférentielle du système : un adjectif antéposé peut presque toujours s'adjoindre au substantif auquel il se rapporte d'une manière postposée, même si cela conduit à un changement de sens (cf. ch. II, § III. 2. 1. ; chV, § IV).

```
      UNE IMAGE DE BONNE QUALIT#E

  658  1                UNE       IMAGE
  660  2  SA                      BONNE ——·——·——·——·— 3 V
  661  3       DE                 QUALIT#E
-------------------------------------------------------------------------------

      POUR LEURS ANCIENS LOGEMENTS

 1532  1  SA                      ANCIENS                2 V
 1533  2       POUR       LEURS  LOGEMENTS
-------------------------------------------------------------------------------

      LE PLANCHER BAS DU DERNIER NIVEAU HABITABLE

 4856  1                LE        PLANCHER
 4857  2                          BAS
 4859  3  SA                      DERNIER                4 V
 4860  4       DU                 NIVEAU
 4861  5                          HABITABLE
-------------------------------------------------------------------------------

      LA NOUVELLE DOCTRINE ADMINISTRATIVE

10315  1  SA                      NOUVELLE               2 V
10316  2                LA        DOCTRINE
10317  3                          ADMINISTRATIVE
-------------------------------------------------------------------------------
```

Nom *de la relation. Il indique le type de la liaison établie et est imprimé sur la même ligne que le mot dépendant*

Note *attribuée à la liaison établie. Elle est censée donner une mesure du degré de de plausibilité ou de validité de la relation ainsi construite. Ceci permettra, comme nous le verrons, d'effectuer, en dernier recours, des choix en cas d'ambiguïté persistante. Les critères qui permettent l'évaluation de cette note seront vus plus loin (cf. ch.V, § IV. 2.).*

Rang *du mot plein auquel se rapporte le mot plein situé sur la même ligne. Ainsi dans la première chaîne nominale, l'adjectif antéposé* ***"bonne"*** *est rapporté au mot de rang 3 dans la chaîne, et par conséquent au mot* ***"qualité".*** ***"V"*** *est la note attribuée à cette liaison.*
En cas d'ambiguïté, plusieurs couples ***(Rang, note)*** *figureront alors sur une même ligne pour indiquer les différents mots auxquels se rapporte le mot associé à ces couples (cf. figures suivantes)*

fig. 27. Reconnaissance des relations Substantif-Adjectif Antéposé
(nous complétons sur cette figure, la description des conventions d'écriture que nous avons choisies pour présenter les résultats obtenus sur le plan informatique).

Exemple :

fausse déclaration	*déclaration fausse*
petit garçon	*garçon petit*
haute fréquence	*fréquence haute*

- Le tableau suivant donne à titre indicatif, les performances du filtre utilisé, mesurées sur un texte d'environ 4 700 mots.

Situation (1)	Silence total (3)	Bruit total (4)
32	0	0

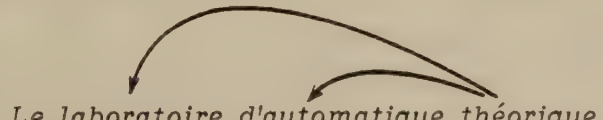
fig. 28. Performances du filtre P_{SA}.

Ainsi, aucun cas de silence ou de bruit n'a été relevé.

b) RLS Substantif - Adjectif Postérieur (SP) :

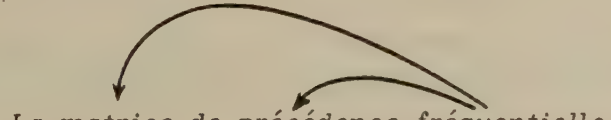
Comme nous l'avons déjà vu dans le chapitre I, cette relation donne généralement lieu à des ambiguïtés dès que, dans une chaîne nominale, plusieurs substantifs précèdent un adjectif postérieur.

Ainsi, dans "Le laboratoire d'automatique théorique" l'accord genre-nombre n'apportant aucune indication, "théorique" devra être rattaché aussi bien à "Laboratoire" qu'à "automatique" :



Le laboratoire d'automatique théorique

Nous avons de même :



La matrice de précédence fréquentielle

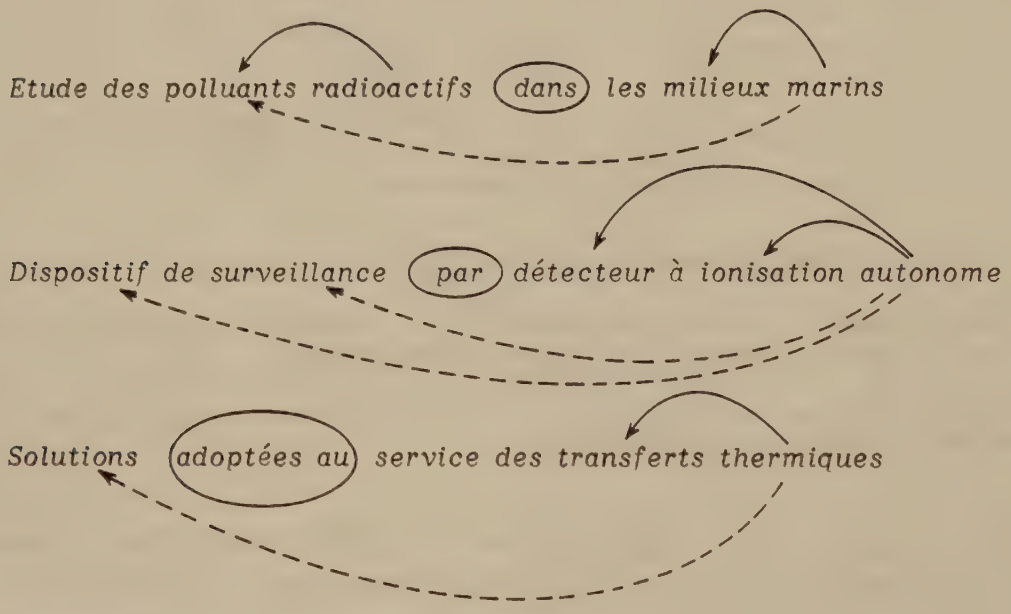
Certaines options peuvent cependant limiter la portée d'investigation lors de la recherche des substantifs susceptibles de régir l'adjectif postérieur, et, partant de réduire les taux d'ambiguïté.

Nous avons en particulier formulé la règle suivante :

La flèche partant d'un adjectif postérieur ne peut enjamber :
- Les prépositions autres que "d', de, du, des, à, au, aux, en" ;
- Les structures "participe passé - préposition"

qui, autrement dit, constituent des "barrières" spécifiques pour la relation SP (cf. ch. III, § II. 3.).

Cette règle, prise en compte par le filtre P^{Cont}_{SP}, permet alors d'éviter la construction de certaines liaisons, telles celles en pointillé dans les exemples suivants :



Remarques :

Les exemples que nous considérons sont choisis pour illustrer les cas où l'accord en genre et en nombre n'apporte aucune indication pour réduire ou résoudre les ambiguïtés inhérentes aux relations SP.

Il est d'ailleurs intéressant de noter que c'est même l'inverse qui se produit. Le deuxième exemple offre ainsi un cas où la détermination du genre de l'adjectif postposé "autonome" demande précisément que l'on ait reconnu le substantif auquel il s'applique.

La règle énoncée ci-dessus n'est cependant maintenue que dans la mesure où elle n'engendre pas de silence susceptible de se traduire par un trop grand nombre de mauvaises reconnaissances (bruit total).

Sur environ 2 500 mots, nous n'avons précisément relevé aucun cas de silence ou bruit total consécutif à l'application de cette règle (cf. fig. 30).

Le filtre syntaxique de reconnaissance des relations SP potentielles se présente donc sous la forme suivante :

$$P_{SP} = F^{S}\ P_{SP}^{Cont}\ F^{AdjP}$$

où P_{SP}^{Cont} est le filtre syntaxique qui permet de prendre en compte les cas de non-contiguïté, fréquents cette fois, et, notamment, la règle que nous avons formulée ci-dessus.

La figure 29 présente des exemples, extraits de textes scientifiques et juridiques, de reconnaissance de relations Substantifs-Adjectifs postposés potentielles.

L'application de la règle énoncée plus haut a permis notamment d'éliminer toute ambiguïté dans les chaînes nominales 1 et 4. Dans la chaîne nominale 3 elle a permis d'éliminer deux liaisons sans toutefois résoudre totalement l'ambiguïté présente. Nous dirons qu'elle a permis de réduire l'ambiguïté puisque deux liaisons potentielles subsistent au lieu de quatre.

Les chaînes nominales 2, 5 et 6 offrent des exemples de reconnaissance où des adjectifs postposés sont coordonnés ou juxtaposés.

Il est clair enfin que toute indication apportée par l'accord en genre et en nombre a été parallèlement prise en compte (cf. exemples 2, 4 et 6 de la fig. 29).

Le tableau suivant donne à titre indicatif les performances du filtre P_{SP} mesurées sur un texte d'environ 2 500 mots.

Situations 1 ou 2	Silence total (3)	Bruit total (4)
126	0	0

fig. 30. Performances du filtre P_{SP}

LA CONNAISSANCE DE SON DEVENIR DANS LE MILIEU MARIN

73	1			LA	CONNAISSANCE		
76	2		DE	SON	DEVENIR		
79	3		DANS	LE	MILIEU		
80	4	SP			MARIN	3 V	

ASPECTS BIOLOGIQUES ET PHYSICO-CHIMIQUES DE LA CONTAMINATION RADIOACTIVE D' ESP∂ECES

197	1				ASPECTS		
198	2	SP			BIOLOGIQUES	1 V	
200	3	SP			PHYSICO-CHIMIQUES	1 V	
203	4		DE	LA	CONTAMINATION		
204	5	SP			RADIOACTIVE	4 V	
206	6		D'		ESP∂ECES		

UN DISPOSITIF DE SURVEILLANCE PAR D#ETECTEUR ∂A IONISATION AUTONOME

506	1			UN	DISPOSITIF		
508	2		DE		SURVEILLANCE		
510	3				D#ETECTEUR		
512	4		∂A		IONISATION		
513	5	SP			AUTONOME	4 D	3 D

LES SOLUTIONS ADOPT#EES AU &SERVICE DES &TRANSFERTS &TERMIQUES

622	1			LES	SOLUTIONS		
623	2				ADOPT#EES		
625	3		AU		&SERVICE		
627	4		DES		&TRANSFERTS		
628	5	SP			&TERMIQUES	4 V	

UNE NOTICE DESCRIPTIVE ET ESTIMATIVE DES TRAVAUX PROJET#ES

2304	1			UNE	NOTICE		
2305	2	SP			DESCRIPTIVE	1 V	
2307	3	SP			ESTIMATIVE	1 V	
2309	4		DES		TRAVAUX		
2310	5				PROJET#ES		

AUX LOCATAIRES DES SOCI#ET#ES ANONYMES COOP#ERATIVES D' HABITATIONS ∂A LOYER MOD#ER#E

1398	1		AUX		LOCATAIRES		
1400	2		DES		SOCI#ET#ES		
1401	3	SP			ANONYMES	2 D	1 D
1402	4	SP			COOP#ERATIVES	2 D	1 D
1404	5		D'		HABITATIONS		
1406	6		∂A		LOYER		
1407	7	SP			MOD#ER#E	6 V	

Fig. 29 Reconnaissance des relations Substantif - Adjectif Postposé (SP)

c) RLS Substantif - Substantif (SS_p) :

La détermination de cette relation donne souvent lieu à des ambiguïtés, dès que dans une chaîne nominale, il y a plus de deux substantifs (cf. ch. I, § I).

Afin de réduire cette ambiguïté (sans toutefois nécessairement la résoudre) certaines règles, visant à restreindre la portée d'investigation lors de la recherche des substantifs susceptibles de régir d'autres substantifs, peuvent néanmoins être formulées, tout au moins dans une première étape.

Nous avons en particulier transposé la règle précédente au cas des relations SS_p :

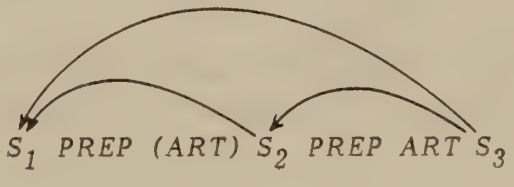
> *Une flèche partant d'un substantif ne peut enjamber :*
> *- les prépositions autres que "à, au, aux, de, du, des, d', en";*
> *- les structures "participe passé - préposition".*

Cette règle contrairement à la règle précédente, s'avèrera cependant, par les silences qu'elles peut occasionner, génératrice d'erreurs dont nous donnons des exemples figure 31.

Une deuxième règle, visant cette fois à éliminer certaines ambiguïtés, a pu également être formulée :

> *Un substantif ne se rapporte qu'au substantif qui le précède immédiatement s'il est non déterminé, sinon il peut se rapporter à tous les substantifs qui le précèdent.*

Ainsi, les structures mettant en jeu trois substantifs S_1, S_2, S_3 seront fléchées comme suit, selon que :

- S_3 est déterminé : 

ou que :

- S_3 n'est pas déterminé :

Cette deuxième règle est cependant, comme la première, susceptible d'engendrer des erreurs dont deux exemples sont donnés figure 32.

A titre indicatif, le tableau suivant donne les performances du filtre P_{SS_p} enregistrées sur un texte d'environ 2 500 mots.

```
      UN G#EN#ERATEUR D' #ELECTRICIT#E @A SOURCE RADIOISOTOPIQUE DE COBALT-60

  798  1                  UN        G#EN#ERATEUR
  800  2  SS    D'                 #ELECTRICIT#E              1 V
  802  3  SS    @A                 SOURCE                     2 V
  803  4  SP                       RADIOISOTOPIQUE            3 D    2 D    1 D
  805  5  SS    DE                 COBALT-60                  3 V
-------------------------------------------------------------------------------------

      LA CONVERSION DE LA CHALEUR EN #ELECTRICIT#E

  813  1                  LA        CONVERSION
  816  2  SS    DE        LA        CHALEUR                    1 V
  818  3  SS    EN                  #ELECTRICIT#E              2 V
-------------------------------------------------------------------------------------
```

fig. 32. "Silences partiels" consécutifs à l'application de la deuxième règle : "SOURCE" se rattache en fait à "GENERATEUR" et "ELECTRICITE" à "CONVERSION".

```
      LA SURFACE B-ATIE DE R#EF#ERENCE VIS#EE @A L' ARTICLE 6

 1898  1                  LA        SURFACE
 1899  2  SP                        B-ATIE                     1 V
 1901  3  VC    DE                  R#EF#ERENCE                2 V
 1902  4  SP                        VIS#EE                     3 V
 1905  5  VC    @A        L'        ARTICLE                    4 V
 1906  6                            6
-------------------------------------------------------------------------------------
      APR@ES AVIS MOTIV#E DU MAIRE

 3505  1        APR@ES              AVIS
 3506  2  SP                        MOTIV#E                    1 V
 3508  3  VC    DU                  MAIRE                      2 V
-------------------------------------------------------------------------------------
```

fig. 31. "Silences" consécutifs à l'application de la première règle : "REFERENCE " se rattache en fait à "SURFACE" et "MAIRE" à "AVIS".

Remarque : Nous verrons que les erreurs qui seront enregistrées seront, pour la plupart, dues à des silences analogues (cf. ch. V, § I. 3.).

Situation 1	Situation 2	Silence total	Bruit total
189	46	10	8
75 %	18 %	4 %	3 %
93 %		4 %	3 %

fig. 33. Performances du filtre P_{SS_p}

Les résultats obtenus semblent ainsi confirmer les deux règles adoptées, dont, rappelons-le, la combinaison permet, en opérant un choix parmi tous les substantifs qui précèdent un substantif donné, de réduire le nombre des candidats auxquels ce substantif est susceptible de se rapporter.

Ces règles, maintenues actuellement, pourront cependant être retirées lorsque l'apprentissage sémantique sera jugé suffisant (cf. ch. V, § IV. 4. 3.). Nous retrouvons ici la dualité, dans l'analyse, des règles syntaxiques/sémantiques.

Nous donnons figure 34 des exemples de reconnaissance de relations Substantif - Substantif.

Nous y trouvons également des relations, nommées **SE**, qui offrent des exemples où le complément déterminatif est "directement" rattaché au substantif qu'il détermine, c'est à dire par simple juxtaposition.

d) RLS Verbe - Sujet (VS), Verbe - complément Direct (VD), Verbe - complément Indirect (VI_p) :

La reconnaissance de ces relations repose sur l'hypothèse selon laquelle, "en général, le participe passé employé sans **avoir** ou **être** - ce qui est le cas dans les chaînes nominales - a le sens passif" (GREVISSE, 62, p. 708)

Nous retenons alors les propriétés suivantes :

LES VARIATIONS DU FACTEUR DE CONCENTRATION DU COBALT-60

```
    3  1                  LES       VARIATIONS
    5  2  SS   DU                   FACTEUR                  1 V
    7  3  SS   DE                   CONCENTRATION            2 V
    9  4  SS   DU                   COBALT-60                3 D   2 D   1 D
```

CHEZ LA MOULE &MYTILUS &EDULIS &L.

```
   12  1       CHEZ      LA        MOULE
   13  2  SE                       &MYTILUS                 1 V
   14  3  SE                       &EDULIS                  1 V
   15  4  SE                       &L.                      1 V
```

L' #ECHANGE PR#EVU ∂A L' ARTICLE 219 DU CODE DE L' URBANISME

```
 1293  1                  L'        #ECHANGE
 1294  2  SP                       PR#EVU                   1 V
 1297  3  VC   ∂A         L'        ARTICLE                  2 V
 1298  4                           219
 1300  5  SS   DU                   CODE                     3 V
 1303  6  SS   DE         L'        URBANISME                5 D   3 D
```

LE MODE D' ALIMENTATION EN EAU

```
 2413  1                  LE        MODE
 2415  2  SS   D'                   ALIMENTATION             1 V
 2417  3  SS   EN                   EAU                      2 V
```

L' EMPLACEMENT DES CANALISATIONS D' #EVACUATION DES EAUX PLUVIALES

```
 2420  1                  L'        EMPLACEMENT
 2422  2  SS   DES                  CANALISATIONS            1 V
 2424  3  SS   D'                   #EVACUATION              2 V
 2426  4  SS   DES                  EAUX                     3 D   2 D   1 D
 2427  5  SP                       PLUVIALES                4 D   2 D
```

D∂ES L' AFFICHAGE ∂A LA MAIRIE D' UN EXTRAIT DE LA D#ECISION

```
 2947  1       D∂ES      L'        AFFICHAGE
 2950  2  SS   ∂A         LA        MAIRIE                   1 V
 2953  3  SS   D'         UN        EXTRAIT                  2 D   1 D
 2956  4  SS   DE         LA        D#ECISION                3 D   2 D   1 D
```

DU PERMIS DE CONSTRUIRE PAR ARR-ET#E DU MINISTRE DE LA RECONSTRUCTION

```
 3483  1       DU                   PERMIS
 3490  2  SS   DE                   CONSTRUIRE               1 V
 3492  3                           ARR-ET#E
 3494  4  SS   DU                   MINISTRE                 3 V
 3497  5  SS   DE         LA        RECONSTRUCTION           4 D   3 D
```

DATE LIMITE

```
11855  1                           DATE
11856  2  SE                       LIMITE                   1 V
```

LE COMMISSIONNAIRE VENDEUR

```
13596  1                  LE        COMMISSIONNAIRE
13597  2  SE                       VENDEUR                  1 V
```

---

fig. 34. Reconnaissance des RLS Substantif - Substantif.

- le substantif auquel il se rapporte est le complément direct, et non le sujet, du verbe dont il provient, il y a de surcroit accord en genre et en nombre ;
- il est susceptible de recevoir les mêmes compléments indirects que le verbe dont il provient ;
- enfin, les compléments d'agent qui peuvent l'accompagner sont les sujets possibles du verbe dont il provient.

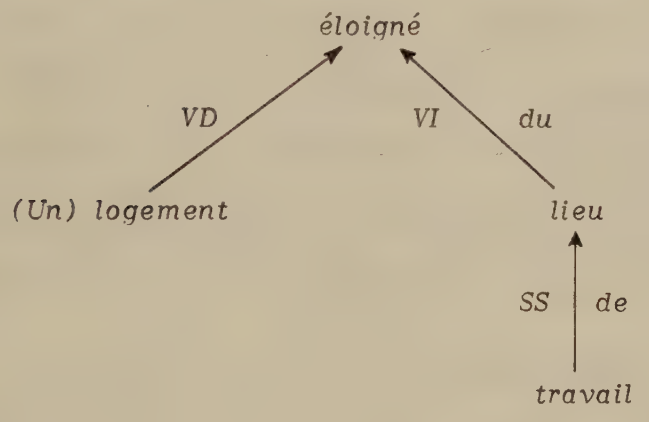
Ainsi, nous associerons aux chaînes nominales :

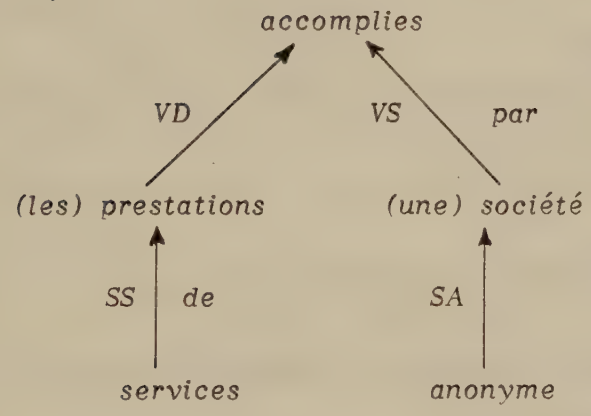
Un logement éloigné du lieu de travail

et

Les prestations de services accomplies par une société anonyme

les représentations suivantes :





Cette hypothèse, vraie en général, est cependant contredite dans des phrases comme :

Une personne passionnée *Une personne repentie*
Une personne décidée *Des oiseaux envolés*
(Grévisse, 62, p.708)

où le participe passé est employé "au sens actif" (62, p.708) , c'est à dire qu'il reflète une construction réfléchie et non transitive.

Nous utiliserons néanmoins les propriétés énoncées ci-dessus, accompagnées de l'option restrictive donnée ci-après, pour définir les règles d'analyse des chaînes nominales incluant un participe passé. Nous verrons en effet que, pratiquement, elles n'ont introduit ni silence ni erreur.

L'option restrictive annoncée est relative à la manière dont les participes passés peuvent recevoir leur complément d'agent, c'est à dire le sujet, dans notre terminologie.

En effet, bien que le complément d'agent puisse être introduit par "de" ou "par" (voire par d'autres prépositions (DUBOIS, 51 ; GROSS, 64)), nous n'avons retenu, du moins dans l'état actuel de l'expérimentation, que la seconde préposition, à savoir "par". Retenir la préposition "de" conduirait en effet, eu égard aux multiples rôles qu'elle peut jouer, à subdiviser les verbes en deux classes, selon qu'ils acceptent un agent introduit par "de" ou non (JAYEZ, 70, p. 135) .

Ajoutons enfin qu'aucune règle restrictive, relative aux cas de non contiguïté, n'a été retenue quant aux relations Verbe - Sujet ou Verbe - complément Indirect. La relation Verbe - complément Direct a par contre hérité de la règle formulée pour les relations SP (cf. § IV. 2. 1. 2. b) transposée :

> *La liaison établie entre un substantitf, candidat à être complément direct, et un participe passé ne peut enjamber*
> *- les prépositions autres que "à", "au", "aux", "de", "du", "des", "d'", "en" ;*
> *- la structure "participe passé - préposition".*

C'est d'ailleurs le même filtre de reconnaissance des relations SP qui est utilisé pour reconnaître les relations VD.

Les figures 35 et 36 offrent respectivement des exemples de reconnaissance de relations Verbe - Sujet et Verbe - complément Direct d'une part, et Verbe - complément Direct, Verbe - complément Indirect d'autre part.

A titre indicatif, nous donnons dans le tableau suivant, les performances des procédures de reconnaissance des relations VS, VD et VI_p, mesurées sur un texte d'environ 5000 mots représentant 140 phrases.

```
      LA D#ECISION D' AGR#EMENT PR#EVUE PAR LE D#ECRET N. 58-1460

2640  1                    LA        D#ECISION
2642  2          D'                  AGR#EMENT
2643  3  SP                          PR#EVUE                    1 V
2646  4  VS                LE        D#ECRET                    3 V
2647  5                              N.
2648  6                              58-1460
-------------------------------------------------------------------------------

      LES EXEMPTIONS PR#EVUES PAR LE PR#ESENT ARR-ET#E

2910  1                    LES       EXEMPTIONS
2911  2  SP                          PR#EVUES                   1 V
2914  3                              PR#ESENT
2915  4  VS                LE        ARR-ET#E                   2 V
-------------------------------------------------------------------------------

      LES TRAVAUX NON EXEMPT#ES PAR L' ARTICLE 1ER

3224  1                    LES       TRAVAUX
3225  2                              NON
3226  3  SP                          EXEMPT#ES                  1 V
3229  4  VS                L'        ARTICLE                    3 V
3230  5                              1ER
-------------------------------------------------------------------------------

      LES GROUPES DE LOGEMENTS DESSERVIS PAR UN M-EME ESCALIER

4552  1                    LES       GROUPES
4554  2             DE               LOGEMENTS
4555  3  SP                          DESSERVIS                  2 D   1 D
4558  4                              M-EME
4559  5  VS                UN        ESCALIER                   3 V
-------------------------------------------------------------------------------

      SAUF D#EROGATION ACCORD#EE PAR LE MINISTRE DE LA CONSTRUCTION

5309  1          SAUF                D#EROGATION
5310  2  SP                          ACCORD#EE                  1 V
5313  3  VS                LE        MINISTRE                   2 V
5316  4          DE        LA        CONSTRUCTION
-------------------------------------------------------------------------------

      @A L' ENSEMBLE DES ACTIVIT#ES LIB#ERALES EXERC#EES PAR DES SOCI#ET#ES ANONYMES

10452 1          @A        L'        ENSEMBLE
10454 2          DES                 ACTIVIT#ES
10455 3                              LIB#ERALES
10456 4  SP                         EXERC#EES                  2 V
10459 5  VS                DES       SOCI#ET#ES                 4 V
10460 6                              ANONYMES
-------------------------------------------------------------------------------
```

fig. 35. Reconnaissance des relations :

- Verbe - Sujet (VS)
- Verbe - complément Direct (notées SP)

DANS LES CONDITIONS D' OCCUPATION INSUFFISANTE D#EFINIES ∂A L' ARTICLE 2

1334	1		DANS	LES	CONDITIONS			
1336	2	SS	D'		OCCUPATION	1	V	
1337	3	SP			INSUFFISANTE	2	V	
1338	4	SP			D#EFINIES	1	V	
1341	5	VC	∂A	L'	ARTICLE	4	V	
1342	6				2			

UN LOGEMENT SITU#E SUR LE TERRITOIRE DE LA M-EME COMMUNE

1347	1			UN	LOGEMENT			
1348	2	SP			SITU#E	1	V	
1351	3	VC	SUR	LE	TERRITOIRE	2	V	
1354	4	SA			M-EME	5	V	
1355	5	SS	DE	LA	COMMUNE	3	V	

LE DOSSIER JOINT ∂A LA DEMANDE DE PERMIS DE CONSTRUIRE

2046	1			LE	DOSSIER			
2047	2	SP			JOINT	1	V	
2050	3	VC	∂A	LA	DEMANDE	2	V	
2052	4	SS	DE		PERMIS	3	V	
2054	5	SS	DE		CONSTRUIRE	4	V	

POUR LES CONSTRUCTIONS #EDIFI#EES SUR LE TERRITOIRE DE LA VILLE DE &PARIS

3024	1		POUR	LES	CONSTRUCTIONS			
3025	2	SP			#EDIFI#EES	1	V	
3028	3	VC	SUR	LE	TERRITOIRE	2	V	
3031	4	SS	DE	LA	VILLE	3	V	
3033	5	SS	DE		&PARIS	4	V	

UN LOGEMENT TROP #ELOIGN#E DU LIEU DE TRAVAIL DU CHEF DE FAMILLE

3610	1			UN	LOGEMENT				
3611	2	PO			TROP	3	V		
3612	3	SP			#ELOIGN#E	1	V		
3614	4	VC	DU		LIEU	3	V		
3616	5	SS	DE		TRAVAIL	4	V		
3618	6	SS	DU		CHEF	5	D	4	D
3620	7	SS	DE		FAMILLE	6	V		

PAR LES APPAREILS DE COMBUSTION SITU#ES DANS LES LOCAUX D' HABITATION

4087	1			LES	APPAREILS			
4089	2	SS	DE		COMBUSTION	1	V	
4090	3	SP			SITU#ES	1	V	
4093	4	VC	DANS	LES	LOCAUX	3	V	
4095	5	SS	D'		HABITATION	4	V	

fig. 36. Reconnaissance des relations :

- Verbe - complément Direct (Elles sont notées SP)
- Verbe - complément Indirect (Elles sont notées VC)

VS			VD			VI_p		
Recon. cor.	Sil. total	Bruit	Recon. cor.	Sil. total	Bruit	Recon. cor.	Sil. total	Bruit
15	0	0	69	0	0	19	0	3

fig. 37.

Ainsi, nous n'avons relevé que trois cas de bruit total, consécutifs d'ailleurs à la non reconnaissance de certaines relations SS_p (cf. fig. 31), c'est à dire donc, à l'application des diverses restrictions que nous avons formulées jusqu'ici. Ce sont ces cas qui donneront le "gros" des erreurs constatées lors de l'acquisition des RLS syntagmatiques (cf. ch. V, fig. 88 et 99).

e) RLS Verbe - Adverbe (VA), Adjectif - Adverbe (AA)

La similitude dans la construction syntaxique de ces deux types de liaisons, la parenté sémantique effective, quelquefois, des relations qu'elles renferment nous ont amené à n'élaborer qu'un seul filtre syntaxique pour les reconnaître.

Les éléments à relier sont en effet presque toujours contigus. La détermination des relations de dépendance VA et AA est dès lors simple et ne conduit à aucune sorte d'ambiguïté.

La figure 38 offre des exemples de reconnaissance de relations VA.

f) RLS Adjectif - Complément d'adjectif (AC_p)

La rareté de cette relation nous a incité, pour réduire certaines ambiguïtés, à négliger, du moins dans une première étape, sa construction.

Cette option constitue cependant une deuxième source de silences dont nous verrons qu'ils sont à leur tour, parfois générateurs d'erreurs (cf. ch. V, § 1.3 fig. 88).

La figure 39 offre ainsi des cas de silence dont certains conduisent à des erreurs.

```
      LES VARIATIONS DE CONTAMINATION PROVOQU#EES IN SITU PAR &RU

  445  1                  LES       VARIATIONS
  447  2  SS   DE                   CONTAMINATION              1 V
  448  3  SP                       PROVOQU#EES                1 V
  449  4  PQ                       IN SITU                    3 V
  451  5  VS                       &RU                        3 V
-------------------------------------------------------------------------------

      ET PAR DES OUVERTURES D' ENTR#EE D' AIR CONVENABLEMENT DISPOS#EES

 3812  1                  DES       OUVERTURES
 3814  2  SS   D'                   ENTR#EE                    1 V
 3816  3  SS   D'                   AIR                        2 V
 3817  4  PO                       CONVENABLEMENT             5 V
 3818  5  SP                       DISPOS#EES                 1 V
-------------------------------------------------------------------------------

      @A DES FOURNITURES PROVISOIREMENT AVANC#EES PAR UNE AUTRE ENTREPRISE

11563  1       @A        DES       FOURNITURES
11564  2  PO                       PROVISOIREMENT             3 V
11565  3  SP                       AVANC#EES                  1 V
11568  4  SA                       AUTRE                      5 V
11569  5  VS                 UNE   ENTREPRISE                 3 V
-------------------------------------------------------------------------------
```

fig. 38. Reconnaissance des relations Verbe - Adverbe (VA)

Elles sont en fait notées PO ou PQ selon que l'adverbe est antérieur ou postérieur au participe passé qu'il modifie.

```
      DES MOD@ELES M#ETABOLIQUES UTILISABLES DANS L' INDUSTRIE NUCL#EAIRE

  139  1                  DES       MOD@ELES
  140  2  SP                       M#ETABOLIQUES              1 V
  141  3  SP                       UTILISABLES                1 V
  144  4       DANS       L'       INDUSTRIE
  145  5  SP                       NUCL#EAIRE                 4 V
-------------------------------------------------------------------------------

      LA NOTE AFF#ERENTE @A CE FOYER

 3623  1                  LA        NOTE
 3624  2  SP                       AFF#ERENTE                 1 V
 3627  3  SS   @A          CE       FOYER                      1 V
-------------------------------------------------------------------------------

      LES QUANTIT#ES D' AIR N#ECESSAIRES @A LEUR BON FONCTIONNEMENT

 4103  1                  LES       QUANTIT#ES
 4105  2  SS   D'                   AIR                        1 V
 4106  3  SP                       N#ECESSAIRES               1 V
 4109  4  SA                       BON                        5 V
 4110  5  SS   @A          LEUR     FONCTIONNEMENT             2 D   1 D
-------------------------------------------------------------------------------
```

fig. 39. Silences "volontaires" dûs à la non reconnaissance des relations Adjectif - Complément d'Adjectif (AC_p).

Notons, à titre indicatif, que nous avons relevé, sur un texte d'environ 5000 mots, 6 cas de silence introduisant des erreurs sur 70. Autrement dit, pour 70 adjectifs suivis d'un groupe prépositionnel, on n'en a trouvé que 6 qui prenaient le groupe prépositionnel pour complément.

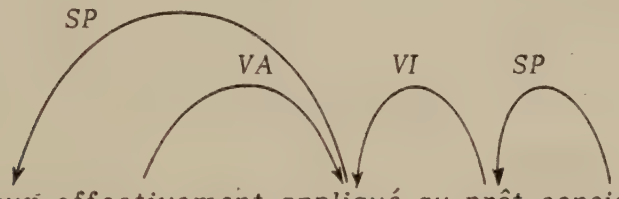
A titre récapitulatif nous donnons figure 40 des exemples d'analyse complète de chaînes nominales dont certaines sont, ainsi que nous pouvons le voir, particulièrement longues.

IV. 2. 1. 3. Eléments libres :

L'analyse dépendancielle de la chaîne nominale ainsi effectuée nous conduit donc à lui associer un ou plusieurs graphes de dépendance dont les noeuds sont les mots pleins de la chaîne nominale et les arcs les relations sémantiques qui s'établissent entre ceux-ci, matérialisés ou non par les mots vides.

Dans cette représentation il subsiste toujours au moins un mot qui demeure non rattaché. Il correspond très souvent au premier mot plein de la chaîne nominale.

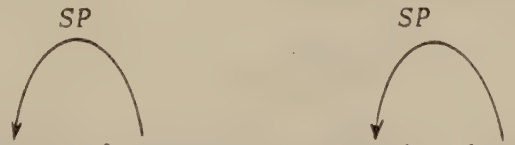
Ainsi, dans la chaîne nominale suivante (cf. fig. 40, exemple 4) :



ex_1 : *Le taux effectivement appliqué au prêt considéré*

qui est fléchée comme indiqué, le mot "taux" n'est rattaché à aucun autre mot; (sur la figure 40 il n'est suivi d'aucune information "rang-note").

De même, dans



ex_2 : ... *porter / un jugement ferme sur une question donnée*

"jugement" et "question" ne sont rattachés à aucun autre mot gouverneur.

Nous dirons que ces mots sont libres ou indépendants.

Inversement, nous dirons que les autres mots (tels que "prêt" dans ex_1 ou "ferme" dans ex_2) sont liés ou dépendants.

LA DISTRIBUTION DE L' #ENERGIE #ELECTRIQUE PAR LIGNES SOUTERRAINES

6067	1			LA	DISTRIBUTION			
6070	2	SS	DE	L'	#ENERGIE	1 V		
6071	3	SP			#ELECTRIQUE	2 D	1 D	
6073	4				LIGNES			
6074	5	SP			SOUTERRAINES	4 V		

LA CONSTRUCTION DES USINES PRODUCTRICES D' #ENERGIE #ELECTRIQUE

6084	1			LA	CONSTRUCTION			
6086	2	SS	DES		USINES	1 V		
6087	3	SP			PRODUCTRICES	2 V		
6089	4	SS	D'		#ENERGIE	2 V		
6090	5	SP			#ELECTRIQUE	4 D	1 D	

L' INSTRUCTION DE LA DEMANDE DE PERMIS DE CONSTRUIRE PAR LE DIRECTEUR D#EPARTEMENTAL DE LA CONSTRUCTION

6544	1			L'	INSTRUCTION			
6547	2	SS	DE	LA	DEMANDE	1 V		
6549	3	SS	DE		PERMIS	2 V		
6551	4	SS	DE		CONSTRUIRE	3 V		
6554	5			LE	DIRECTEUR			
6555	6	SP			D#EPARTEMENTAL	5 V		
6558	7	SS	DE	LA	CONSTRUCTION	5 V		

LE TAUX EFFECTIVEMENT APPLIQU#E AU PR-ET CONSID#ER#E

8493	1			LE	TAUX			
8494	2	PO			EFFECTIVEMENT	3 V		
8495	3	SP			APPLIQU#E	1 V		
8497	4	VC	AU		PR-ET	3 V		
8498	5	SP			CONSID#ER#E	4 V		

L' APPLICATION DU PRINCIPE DE COMMERCIALIT#E DES PRESTATIONS DE SERVICES R#EALIS#EES PAR LES SOCI#ET#ES ANONYMES

10611	1			L'	APPLICATION			
10613	2	SS	DU		PRINCIPE	1 V		
10615	3	SS	DE		COMMERCIALIT#E	2 V		
10617	4	SS	DES		PRESTATIONS	3 D	2 D	1 D
10619	5	SS	DE		SERVICES	4 V		
10620	6	SP			R#EALIS#EES	4 V		
10623	7	VS		LES	SOCI#ET#ES	6 V		
10624	8	SP			ANONYMES	7 V		

LES LIVRAISONS DE MARCHANDISES FAITES EN RESTITUTION D' UN PR-ET DE CONSOMMATION CONSENTI DANS LES CONDITIONS D#EFINIES AUX ARTICLES 1892

12847	1			LES	LIVRAISONS			
12849	2	SS	DE		MARCHANDISES	1 V		
12850	3	SP			FAITES	2 D	1 D	
12852	4	VC	EN		RESTITUTION	3 V		
12855	5	SS	D'	UN	PR-ET	4 V		
12857	6	SS	DE		CONSOMMATION	5 V		
12858	7	SP			CONSENTI	5 V		
12861	8	VC	DANS	LES	CONDITIONS	7 V		
12862	9	SP			D#EFINIES	8 V		
12864	10	VC	AUX		ARTICLES	9 V		
12865	11	SP			1892	10 V		

---

fig. 40. Construction des relations de dépendance potentielles internes aux chaînes nominales.

Cette information attribuée à chaque mot plein de la chaîne nominale sera, comme nous le verrons, utile dans la suite du traitement, lors de la reconnaissance des relations hétérosyntagmatiques (§ IV. 2. 3.).

Nous retrouvons ici, encore une fois, la construction , au fur et à mesure que l'on avance dans l'analyse, de nouvelles informations à prendre en compte dans les étapes ultérieures du traitement.

IV. 2. 2. Analyse de la chaîne verbale :

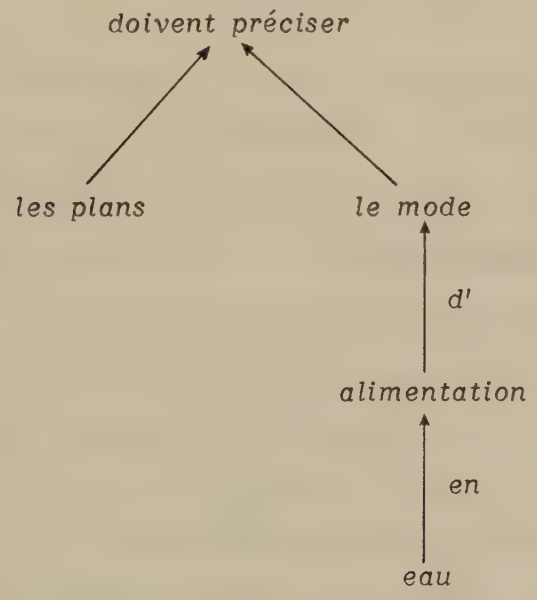
Nous allons décrire dans ce paragraphe les grandes lignes des options limitatives, et par conséquent simplificatrices, que nous avons retenues pour l'analyse de la chaîne verbale.

Tout d'abord nous n'avons pas jugé utile, du moins dans une première étape, d'obtenir une représentation sous forme de dépendance de la chaîne verbale.

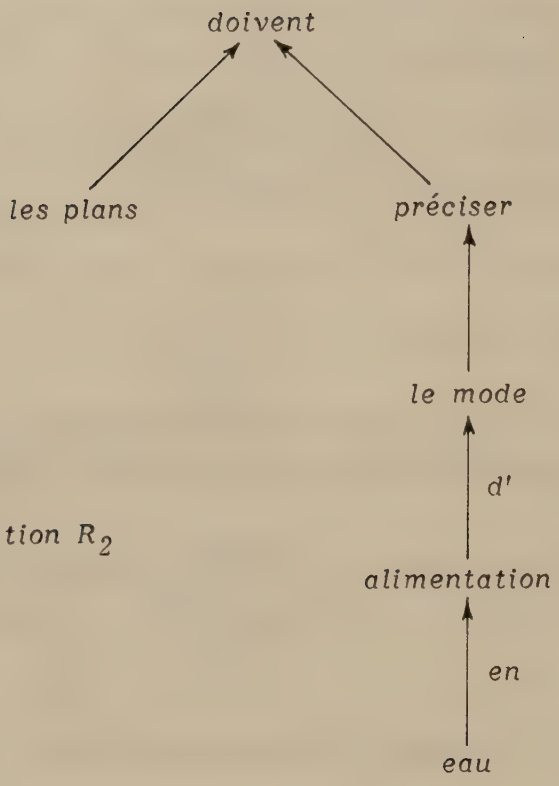
Ainsi, la phrase :

S_1 : *Les plans doivent préciser le mode d'alimentation en eau.*

recevra la représentation R_1 et non la représentation plus détaillée R_2 (cf TESNIERE, 108, pp. 47, 107, 398) :



Représentation R_1



Représentation R_2

La phrase :

S_2 : *Les plans précisent le mode d'alimentation en eau*

aura donc le même graphe dépendanciel, à savoir le graphe R_1.

Nous chercherons néanmoins à reconnaître, dans la chaîne verbale, ce que nous appellerons son **centre** et qui correspondra au verbe, dépouillé des divers auxiliaires ou semi-auxiliaires (de temps, de mode, etc.) qui peuvent l'accompagner ainsi que des adverbes éventuels.

Ainsi, dans les chaînes suivantes, le centre verbal sera :

imposer	*dans*	*pourra être imposé*
déposer	"	*auront déposé*
assurer	"	*ont pour objet d'assurer*
équiper	"	*est considéré comme équipé*
fermer	"	*sont équipés de manière à fermer hermétiquement*
fixer	"	*doivent alors être solidairement fixés*
préciser	"	*doivent préciser*
localiser	"	*permettant de localiser*
connaître	"	*doit faire connaître*

Par suite, seul sera pris en compte le centre verbal dans l'économie des diverses relations sémantiques faisant intervenir la chaîne verbale.

Cette option sera toutefois source de silences.

En effet, outre les relations LS Verbe - Adverbe (VA) qui seront ainsi sacrifiées, d'autres relations LS (certaines relations Verbe - Sujet (VS), notamment) ne pourront être reconnues. Nous visons ici en particulier les chaînes verbales faisant intervenir des verbes factitifs et qui, eu égard aux diverses relations dont elles sont le siège, pourraient être considérées comme présentant deux centres verbaux (*).

Ainsi, dans :

S_3 : Le professeur fait réciter la leçon à l'élève

la chaîne verbale "fait réciter" présente deux éléments verbaux (faire et réciter) qui ont leurs propres sujet ou compléments (cf. fig. 41).

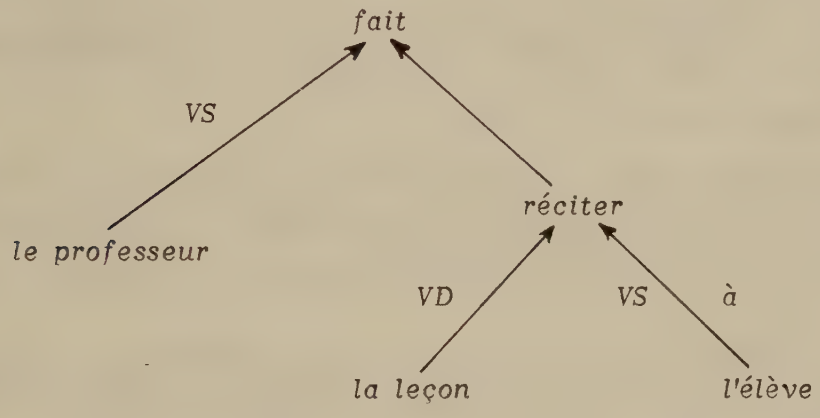


fig. 41

Le sujet grammatical de la forme factitive ne correspond pas en effet à l'agent de l'action.

Or ceci n'est pas uniquement le fait des verbes factitifs (faire, laisser). D'autres chaînes verbales peuvent présenter des interprétations analogues.

C'est par exemple le cas dans la phrase :

S_4 : L'organisateur de la conférence voit arriver les participants.

(*) voire même trois ou plus. C'est le cas dans les chaînes telles que "faire faire écrire" par exemple. (voir TESNIERE, 108, chap. 107 à 112) .

qui reçoit la description structurale donnée figure 42, et où l'on voit figurer tout comme pour la phrase S_3 précédente, deux relations Verbe - Sujet (VS).

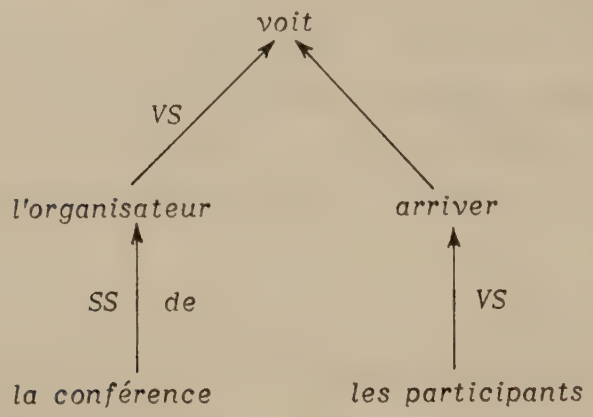


fig. 42

A cet égard, les verbes tels que "regarder", "voir", "sentir", "entendre", etc. (108, p. 433; 62, p. 1035) pourraient donc être adjoints aux verbes factitifs traditionnels pour constituer une classe particulière, utile pour la reconnaissance des diverses relations VS, VD ou VI_p.

Nous verrons néanmoins, que les chaînes verbales mettant en jeu de tels verbes sont rares dans le corpus (cf. fig. 43). Les erreurs qui pourraient survenir, du fait de l'option que nous avons adoptée, seront par conséquent relativement peu nombreuses (cf. ch. V, fig. 89).

Les auxiliaires de temps ou de mode étant toujours antéposés au verbe, la reconnaissance de ce dernier est dès lors relativement simple : c'est le dernier élément verbal de la chaîne ainsi que les exemples précédents le montrent.

Si sur le plan sémantique il y a lieu de reconnaître le centre verbal qui seul intervient dans les relations LS, sur le plan syntaxique ce sont, et le centre verbal et les auxiliaires, lorsqu'ils sont présents, qui renferment les informations nécessaires à la détermination ultérieure des diverses relations LS (VS, VD ou VI_p).

Il y a lieu en effet de déterminer la voix (active ou passive) de la chaîne verbale d'une part, et de lui attribuer les informations relatives aux nombre, personne et éventuellement genre, d'autre part (*).

(*) Les informations relatives aux temps et aux modes n'intervenant pas, ou du moins que très rarement, dans la recherche des relations de dépendance potentielles, nous ne nous y sommes pas intéressés. Voir cependant § IV.2.3.1.a.

Le décryptage de ces informations constituera le deuxième volet de ce que nous avons appelé analyse de la chaîne verbale, le premier étant la détermination de son centre.

La reconnaissance de la négation constitue enfin le troisième volet de cette analyse.

a) Type de la chaîne verbale :

La détermination ultérieure des diverses relations de dépendance potentielles (VS, VD ou VI_p) nous a amenés en fait à effectuer, non pas une simple dichotomie, "forme active" "forme passive", des chaînes verbales, mais une catégorisation plus complexe de celles-ci, qui, en réalité, reviendra pour l'essentiel à effectuer une partition des formes actives obtenues.

La présente étape de l'analyse consiste donc à attribuer à la chaîne verbale ce que nous appellerons un type, dont le but est de conditionner ultérieurement la recherche des différents sujets ou compléments potentiels.

Nous nous bornerons ici à donner, à titre indicatif, les différents types que nous avons retenus et dont nous verrons l'utilité dans les paragraphes suivants.

- **Type O :** C'est la catégorie qui regroupe les verbes dits d'état ou copules et dont le rôle est de rattacher l'attribut au sujet. **Ex. :** Le germanium est un semi-conducteur.

- **Type C :** C'est la catégorie qui regroupe la forme active des verbes comme "rendre", "juger", "nommer", "élire", etc. et dont la particularité est de rattacher l'attribut au complément direct (108, p.161 ; 62, p. 148 ; 111, p. 69) **Ex. :** Le magistrat juge coupable le délinquant.

- **Type P :** C'est la catégorie qui regroupe toutes les formes passives.
  Ex. : Le délinquant est jugé coupable.
  Les exonérations sont prévues par le présent arrêté.

- **Type S :** C'est la catégorie qui regroupe les formes actives des verbes autres que ceux entrant dans les types C ou O.
  Ex. : Le conseil siège au ministère de la construction.

Ces catégories feront en fait l'objet d'une autre sous-catégorisation que nous verrons plus loin (cf. ch. III, § IV. 2. 3. 1. a, fig. 45)

b) Valeurs du nombre, personne, genre :

Il s'agit ici de déterminer les valeurs du nombre, de la personne et éventuellement du genre lorsque, notamment, la chaîne verbale est complexe.

Alors que la valeur de la personne est toujours supportée par le seul premier élément verbal de la chaîne, les valeurs du nombre et du genre sont, selon les cas, supportés soit seulement par le (ou les) premier élément verbal de la chaîne (c'est le cas des formes actives), soit conjointement par le premier élément verbal (auxiliaire ou semi-auxiliaire) et le dernier élément verbal (centre) de la chaîne (c'est le cas des formes passives).

Analyse de la négation :

La négation relève des relations satellites. Elle n'a cependant pas fait l'objet d'une relation mais d'un calcul qui a eu pour but d'attribuer un signe positif ou négatif (+ ou -) à la chaîne verbale, indiquant ainsi les deux éventualités.

Par ailleurs, nous n'avons retenu que les cas relativement simples, mais répandus, où la négation *"se fait en deux temps au moyen de deux marquants dont le premier,* ***ne*** *, est le discordantiel et le deuxième,* ***pas*** *, le forclusif."* (108, p. 224)

Nous avons en fait retenu comme autres forclusifs "point", mais aussi "plus", "jamais", etc.. Le limitatif "que " n'a par contre pas été retenu.

Cette analyse simple assimile donc la négation partielle à la négation totale (111, p. 389-394) .

Il convient cependant de noter que la simplification "positif-négatif", jointe à l'autre approximation : n'avoir qu'un centre pour la chaîne verbale, induit une confusion potentielle ; par exemple :

+	-
peut entendre	*ne peut pas entendre* *peut ne pas entendre*

Mais cette analyse reste, nous semble-t-il, suffisante, du moins dans la perspective de l'analyse dépendancielle et l'acquisition des RLS (voir remarque plus loin).

L'analyse de la chaîne verbale ainsi définie ne constitue donc en réalité qu'une pré-étape pour la recherche des diverses relations VS, VD ou VI_p.

Elle consiste en résumé, pour l'essentiel :

- d'une part, à reconnaître l'élément central de la chaîne verbale qui interviendra en définitive dans les RLS ;
- d'autre part, à attribuer les informations morpho-syntaxiques qui conditionnent les recherches des différents sujets ou compléments potentiels.

La figure 43, ainsi que la figure 17 offrent précisément des exemples de reconnaissance de l'élément central et du type de chaînes verbales, obtenues elles-mêmes automatiquement lors d'une précédente étape (cf. ch. III, § III. 2.)

Les chaînes verbales marquées d'une astérisque offrent en particulier des exemples de chaînes présentant en fait deux centres verbaux (cf. fig. 43).

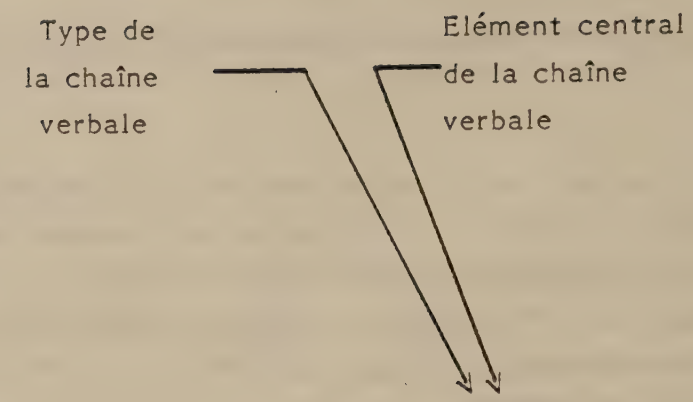
Le tableau suivant donne à titre indicatif les proportions de ces cas comptabilisés sur un texte d'environ 5000 mots, comptant environ 260 chaînes verbales.

Nbre total de chaînes verbales	Chaînes verbales présentant un seul centre	Chaînes verbales présentant deux centres	
		fractitives	Autres
257	245 / 95,3%	5 / 2%	7 / 2,7 %

fig. 44

Indiquons également que ces chaînes verbales se répartissent comme suit quant à leur type :

. 159 sont de type S
. 86 P
. 11 O
et . 1 est de type C



CHAINES VERBALES RECONNUES

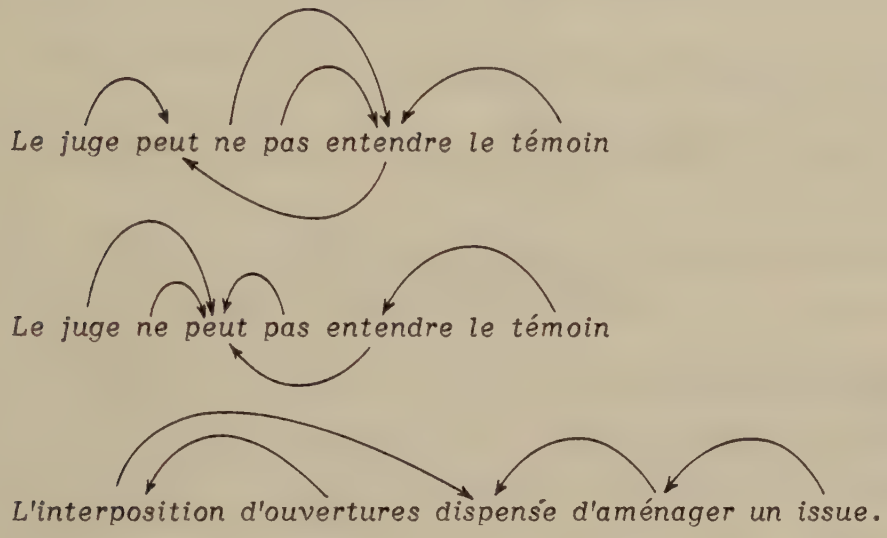
ENTRE EN EFFET	S ENTRE
S' ASSOCIE	S ASSOCIE
RENDENT N#ECESSAIRE	C N#ECESSAIRE
PERMET D' OBTENIR	S OBTENIR
EST	O EST
A FAIT APPARA¬ITRE	S APPARA¬ITRE
CONDUISIRENT ∂A SUPPOSER	S SUPPOSER
ONT #ET#E #ETUDI#EES	P #ETUDI#EES
AFIN D' #ETABLIR	S #ETABLIR
A #ET#E COMPAR#EE	P COMPAR#EE
FAIT VARIER	S VARIER
EST	O EST
NE PEUT ¬ETRE IMPOS#E	P IMPOS#E
EST CONSID#ER#E COMME #EQUIP#E	P #EQUIP#E
N' EST PAS DOT#E	P DOT#E
COMPREND	S COMPREND
COMPORTANT	S COMPORTANT
PERMETTANT DE LOCALISER	S LOCALISER
DOIT FAIRE CONNA¬ITRE	S CONNA¬ITRE
ONT POUR OBJET D' ASSURER EN PERMANENCE	S ASSURER
SOIENT MAINTENUES	P MAINTENUES
OUVRENT	S OUVRENT
SONT #EQUIP#EES DE MANI∂ERE ∂A FERMER HERM#ETIQUEMENT	S FERMER
DOIT ¬ETRE ASSUR#E DE FA$CON PERMANENTE	P ASSUR#E
POURRONT ¬ETRE PR#ECIS#EES	P PR#ECIS#EES
DOIVENT ¬ETRE TELS	O ¬ETRE
NE SOIENT PAS FACILEMENT OBTUR#EES	P OBTUR#EES
PUISSENT ¬ETRE ENTRETENUES	P ENTRETENUES
R#ESULTE	S R#ESULTE
NE DOIVENT OCCASIONNER	S OCCASIONNER
POUVANT COMPROMETTRE	S COMPROMETTRE
NE DOIVENT S' #ECARTER	S #ECARTER
DOIVENT ¬ETRE DE TOUTE MANI∂ERE LIMIT#ES	P LIMIT#ES
DOIVENT ALORS ¬ETRE SOLIDAIREMENT FIX#ES	P FIX#ES

fig. 43. Reconnaissance de l'élément central et du type de la chaîne verbale.

Remarque :

En fait la chaîne verbale devrait, au même titre que la chaîne nominale, recevoir une description dépendancielle. On pourrait ainsi représenter correctement la portée de la négation ; en outre, il serait possible de rattacher un substantif donné au verbe dont il dépend effectivement.

Exemples :



Toutefois, il nous semble que ce genre d'aménagement serait surtout utile pour représenter fidèlement le contenu d'un texte précis, et moins pour ce qui nous intéresse principalement ici, à savoir l'acquisition de relations sémantiques binaires potentielles. Par exemple, du texte "Le juge peut entendre...", il est plus intéressant d'extraire la RLS (juge, entendre, VS) plutôt que (juge, pouvoir, VS).

IV. 2. 3. Construction des relations hétérosyntagmatiques potentielles :

Il s'agit de déterminer maintenant les différentes relations qui peuvent s'établir entre la chaîne verbale, plus précisément son centre, d'une part, et les différents substantifs environnants d'autre part.

Les relations à établir, hétérosyntagmatiques désormais, sont donc en l'occurrence les relations Verbe - Sujet, Verbe - complément Direct et Verbe - complément Indirect.

Nous ne tenterons pas toutefois d'établir toutes les relations potentielles VS, VD ou VI_p de manière exhaustive. Nous ne le pourrions pas.

En effet, la position des sujets ou compléments vis à vis du verbe n'est pas fixe.

Si d'ordinaire le sujet se place avant le verbe (mis à part le cas du passif) et les compléments après ; nombreuses sont également les tournures où cet ordre est inversé (62 ; 111) .

Il en est ainsi dans les phrases suivantes par exemple :

Sur la colline chevauchaient plusieurs cavaliers.
Tant va la cruche à l'eau qu'elle se brise.
Les habitations que proposent les organismes

Etablir toutes les relations potentielles pourrait alors conduire à des configurations d'ambiguïté non justifiées par la nature des textes traités qui présentent, ainsi que nous le verrons, une relative stabilité dans la position des sujets ou des compléments vis à vis du verbe.

Rappelons cependant que nous avons voulu le système "ouvert" et qu'il y a par conséquent toujours possibilité de construire, de proche en proche, toutes les relations potentielles. Notre but est présentement d'illustrer la méthodologie que nous proposons.

Une première option limitative que nous avons adoptée est donc liée aux positions relatives des sujets ou compléments vis à vis de la chaîne verbale.

Nous avons en effet retenu que :

- pour les formes actives, le sujet est toujours placé avant la chaîne verbale et les compléments après ;
- pour les formes passives, les sujet et complément indirect éventuels sont toujours placés après la chaîne verbale et le complément direct avant.

Cette option n'implique évidemment pas que les sujets ou compléments soient contigus à la chaîne verbale.

D'autres options propres à chaque relation seront donc là encore nécessaires (barrières, cf. ch. III, § II. 3.).

Ces options nous amèneront par conséquent à élaborer des filtres de reconnaissance analogues aux filtres syntaxiques précédents (cf. ch. III, § IV. 2.).

IV. 2. 3. 1. Reconnaissance des relations VS potentielles :

a) Dans les formes actives :

Ces formes ont été reconnues lors de l'étape précédente. Elles avaient de surcroît fait d'emlée l'objet d'une première catégorisation : O, C et S ; la catégorie S regroupant la plupart des formes actives reconnues (cf. § IV. 2. 2.).

La reconnaissance des relations Verbe - Sujet potentielles nous amène à élaborer une deuxième sous-catégorisation.

La reconnaissance des sujets "candidats" dépend en effet de la forme du verbe.

Dans les deux phrases suivantes par exemple :

Le ministe de la construction autorise ...
Le plan de la construction comportant ...

alors que dans la première, les informations morpho-syntaxiques permettent de réduire l'ensemble des sujets potentiels au singleton < ministre > , dans la deuxième, "plan" et "construction" restent concurremment candidats.

Dans la phrase suivante par contre :

Pour tout renseignement complémentaire, écrire à ...

le sujet n'est même pas exprimé.

Ces considérations nous ont donc amenés à effectuer une sous-catégorisation qui pourrait correspondre en partie à la classification traditionnelle selon les modes des verbes (GREVISSE, 62, p. 553) .

Nous avons en effet retenue trois sous-catégories :

- la première, notée **MP** (Mode Personnel), correspond aux formes de l'indicatif, du conditionnel et du subjonctif ;
- la deuxième, notée **MIP** (Mode Impersonnel Participe), correspond aux formes participe présent du verbe ;
- la troisième enfin, notée **MII** (Mode Impersonnel Infinitif), regroupe les formes infinitives, mais aussi impératives, du verbe.

Cette dernière catégorisation complète donc le deuxième volet de l'analyse de la chaîne verbale (cf. § IV. 2. 2. a), volet qui peut être résumé comme sur la figure suivante :

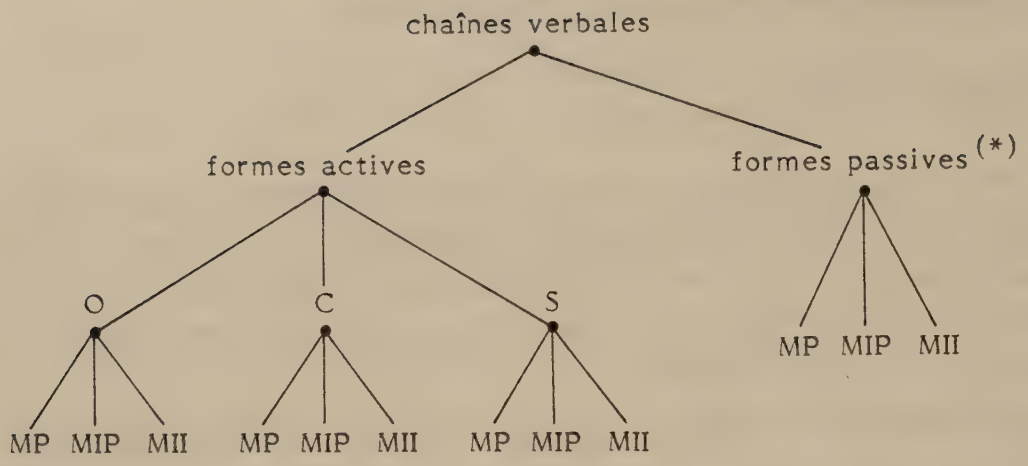


fig. 45.

Pour ce qui concerne la détermination automatique de cette dernière information, notons enfin qu'elle est contenue dans le premier élément verbal de la chaîne.

Dans tous les cas énumérés la reconnaissance des relations VS potentielles consistera donc, dans l'étape présente, à déterminer les différents éléments (Substantifs ou pronoms) susceptibles d'être rattachés, en tant que sujets, à une chaîne verbale donnée.

Dans tous les cas également, la reconnaissance sera fondée sur deux types d'options :

- les premières sont relatives à la limitation du champ d'investigation, c'est à dire aux barrières qui ne pourront être franchies lors de la recherche des sujets potentiels ;
- les secondes ont trait à la détermination des sujets potentiels ou plus précisément aux conditions que ceux-ci doivent remplir.

Afin d'illustrer ces propos, nous donnons les principales options que nous avons retenues pour chacune des catégories ainsi adoptées.

Soulignons cependant à leur sujet, que comme toutes les options précédentes ou à venir, elles pourront être complétées, modifiées ou supprimées selon la qualité des résultats obtenus.

(*) Les formes passives subissent en fait également une sous-catégorisation analogue, selon que le premier élément verbal est à l'indicatif, au conditionnel ou au subjonctif (MP), au participe présent (MIP) ou enfin à l'infinitif (MII). Cette sous-catégorisation interviendra lors de la détermination des relations VD dans les formes passives (voir plus loin § IV 2. 3. 2. b).

1°) Reconnaissance dans les formes MP :

• **Limitation du champ d'investigation :** Nous n'établissons aucune barrière particulière, autrement dit le champ d'investigation est la totalité du début de la phrase.

• Le champ d'investigation étant ainsi délimité, l'élément (substantif ou pronom) susceptible d'être sujet d'une chaîne verbale à l'indicatif, au subjonctif ou au conditionnel doit répondre aux conditions suivantes :

- être libre (cf. ch. III, § IV. 2. 1. 3.).
- lorsqu'il est substantif, ne pas avoir de relation satellite prépositive (cf. ch. III, § IV. 1. et fig. 25).

Ces deux conditions étant satisfaites, c'est de plus l'élément le plus proche géographiquement de la chaîne verbale qui est pris en compte.

Il est clair enfin qu'il y a lieu ici de vérifier l'accord en nombre et en personne.

Ces options permettent de reconnaître le sujet dans un grand nombre de configurations syntaxiques dont nous donnons quelques exemples figure 46.

Remarque :

Les pronoms relatifs "qui", bien que n'apparaissant pas sur la figure (voir les phrases n° 30 et 102), sont néanmoins reconus lors de la présente étape et rattachés aux verbes dont ils sont sujets. Ils deviennent dès lors liés. D'où la reconnaissance correcte du sujet dans les deux phrases 30 et 102 qui présentent des configurations syntaxiques où la distance Sujet - Verbe est particulièrement importante. Leur étude sera complétée dans le paragraphe IV. 2. 4. consacré au traitement de l'anaphore.

Le tableau suivant (fig. 47) donne, à titre indicatif, les taux de bruit ou de silence comptabilisés sur un texte d'environ 5000 mots.

Nature du sujet												Performances globales			
Substantifs				Pronoms				Pr. rel. qui							
Rec.	Silence			Rec.	Silence			Rec.	Silence			Rec.	Silence		
	part.	total			p.	t.			p.	t.			p.	t.	
cor.				cor.				cor.				cor.			
72	1	2		34	0	0		16	0	0		122	1	2	
96 %	1.3 %	2.7 %		100 %	-	-		100 %	-	-		97.6 %	0.8 %	1.6 %	

fig. 47

```
        PHRASE NO :  6   ELLE COMMENCE AU MOT NO:  4851

        $00001300 LES HABITATIONS DONT LE PLANCHER BAS DU DERNIER NIVEAU HABITABLE EST SITU#E @A PLUS DE
        VINGT-CINQ M=ETRES DU SOL DOIVENT COMPORTER UNE OU PLUSIEURS COLONNES S=ECHES DE 65 / 70 MM DONT L'
        INSTALLATION DOIT ¬ETRE ARR¬ET#EE EN ACCORD AVEC LES SERVICES D' INCENDIE ( PR#EFECTURE DE POLICE DE
        &PARIS , INSPECTION D#EPARTEMENTALE DES SERVICES D' INCENDIE DANS LES D#EPARTEMENTS ) .

.....

        DOIVENT COMPORTER                                                    S COMPORTER

  4853  1  VS            LES       HABITATIONS              6 V
  4871  6                          COMPORTER
-----------------------------------------------------------------------------------------

        PHRASE NO : 24   ELLE COMMENCE AU MOT NO:  619

        ON D#ECRIT LES SOLUTIONS ADOPT#EES AU &SERVICE DES &TRANSFERTS &TERMIQUES DU &CENTRE D' &ETUDE
        &NUCL#EAIRE DE &GRENOBLE .

.....

        D#ECRIT                                                              S D#ECRIT
SUJET DE TYPE PRONOMINAL :  ON

-----------------------------------------------------------------------------------------

        PHRASE NO : 30   ELLE COMMENCE AU MOT NO:  5789

        $00001370 LES ORGANISMES D' HABITATIONS @A LOYER MOD#ER#E QUI , @A TITRE EXCEPTIONNEL ,
        JOUENT LE R¬OLE DE MANDATAIRE ADMINISTRATIF DES PERSONNES PHYSIQUES D#ESIREUSES D' ACC#EDER @A LA
        PROPRI#ET#E PEUVENT PERCEVOIR , @A TITRE DE R#EMUN#ERATION , UNE CONTRIBUTION AU PLUS #EGALE @A 1 %
        DU PRIX DE REVIENT DE L' OP#ERATION .

.....

        PEUVENT PERCEVOIR                                                    S PERCEVOIR

  5791  1  VS            LES       ORGANISMES               7 V
  5815  7                          PERCEVOIR
-----------------------------------------------------------------------------------------

        PHRASE NO : 33   ELLE COMMENCE AU MOT NO:  838

        CHAQUE TURBINE ENTRAINE UN ALTERNATEUR .

.....

        ENTRAINE                                                             S ENTRAINE

   839  1  VS            CHAQUE    TURBINE                  2 V
   840  2                          ENTRAINE

-----------------------------------------------------------------------------------------
```

fig. 46. (a). Reconnaissance du sujet dans le cas des formes à l'indicatif, au subjonctif ou au conditionnel (formes MP).

PHRASE NO : 36 ELLE COMMENCE AU MOT NO: 5988

PENDANT LA P*ERIODE O@U LA SOCI*ET*E DE CR*EDIT IMMOBILIER G@ERE LA SOCI*ET*E CIVILE IMMOBILI@ERE , ELLE PER$COIT ANNUELLEMENT UNE R*EMUN*ERATION , POUR FAITS DE GESTION , DANS LA LIMITE DE 0,75 % DU MONTANT INITIAL DU PR~ET PRINCIPAL ET DU MONTANT INITIAL DU PR~ET CONSENTI , LE CAS *ECH*EANT , PAR LA SOCI*ET*E DE CR*EDIT IMMOBILIER @A LA SOCI*ET*E CIVILE POUR L' ACQUISITION DU TERRAIN .

G@ERE S G@ERE

5993	2	VS	LA	SOCI*ET*E	4 V	
5997	4			G@ERE		

PER$COIT ANNUELLEMENT S PER$COIT

SUJET DE TYPE PRONOMINAL : ELLE

PHRASE NO :102 ELLE COMMENCE AU MOT NO: 3404

CES CONSTRUCTIONS QUI FONT APPEL @A DES PRINCIPES NON TRADITIONNELS PR*ESENTENT SOIT DANS L' AGENCEMENT INT*ERIEUR DES PLANS , SOIT DANS LES DIMENSIONS DE BASE DES LOCAUX , SOIT DANS L' *EQUIPEMENT DES PI@ECES PRINCIPALES OU SECONDAIRES , UN CARACT@ERE NOVATEUR PAR RAPPORT AUX USAGES ET @A LA R*EGLEMENTATION EN VIGUEUR .

PR*ESENTENT S PR*ESENTENT

3405	1	VS	CES	CONSTRUCTIONS	3 V	
3413	3			PR*ESENTENT		

PHRASE NO :114 ELLE COMMENCE AU MOT NO: 3761

$00000990 DANS LES PI@ECES PRINCIPALES , LORSQUE TOUTES LES BAIES D' UN LOGEMENT OUVRENT SUR UNE SEULE FA$CADE , DANS LES PI@ECES DONT LES BAIES SONT *EQUIP*EES DE MANI@ERE @A FERMER HERM*ETIQUEMENT , LE RENOUVELLEMENT DE L' AIR DOIT ~ETRE ASSUR*E DE FA$CON PERMANENTE PAR DES OUVERTURES D' *EVACUATION DE L' AIR ET PAR DES OUVERTURES D' ENTR*EE D' AIR CONVENABLEMENT DISPOS*EES ET AM*ENAG*EES , DE SECTION SUFFISANTE , QUI POURRONT ~ETRE PR*ECIS*EES PAR DES INSTRUCTIONS INTERMINIST*ERIELLES .

OUVRENT S OUVRENT

3770	2	VS	LES	BAIES	4 V	
3774	4			OUVRENT		

SONT *EQUIP*EES DE MANI@ERE @A FERMER HERM*ETIQUEMENT S FERMER

3785	6	VS	LES	BAIES	7 V	
3789	7			FERMER		

PHRASE NO : 12 ELLE COMMENCE AU MOT NO: 9676

LES ACHATS EFFECTU*ES PAR UN COMMER$CANT POUR LES BESOINS DE SON COMMERCE SONT PR*ESUM*ES COMMERCIAUX ; L' ACHETEUR NON COMMER$CANT EST PR*ESUM*E ACHETER POUR SON USAGE PERSONNEL .

EST PR*ESUM*E ACHETER S ACHETER

9693	5	VS	L'	ACHETEUR	6 V	
9698	6			ACHETER		COMPL

---

fig. 46. (b).

Les remarques suivantes peuvent être faites :

- En raison de la proximité des éléments de la relation Verbe - Sujet lorsque ce dernier est pronom, aucun bruit ou silence n'a été enregistré sur un total de 50 relations VS mettant en jeu précisément des pronoms (personnels ou relatifs)
- Dans l'autre cas par contre, lorsque le sujet est substantif, il est alors souvent non contigu au verbe auquel il se rapporte. Une erreur et deux silences ont été enregistrés dans ce cas. Nous donnons figure 48 les phrases qui ont donné lieu à ces anomalies.

- La première, relative à l'erreur, présente, ainsi que nous pouvons le voir, une configuration syntaxique où le sujet est situé précisément après le verbe (voir option plus haut). Elle nous suggère également de compléter les options ayant trait aux barrières, en introduisant comme tels les pronoms relatifs compléments (que, dont, etc..). Ces barrières n'auraient cependant d'effet que pour la première chaîne verbale qui les suit, c'est à dire celle dont ils sont compléments.
- La seconde phrase, siège d'un silence total, offre un exemple de non-reconnaissance due à une ambiguïté sur le premier mot "des" non correctement résolue lors d'un traitement précédent. Ceci apporte donc une illustration de l'utilité qu'il y a à imbriquer les traitements des différents niveaux. En particulier, il y a lieu ici de faire un retour en arrière et d'utiliser en "rétro-action" la présente information, à savoir la non-reconnaissance du sujet, pour résoudre correctement le mot "des" (cf. introduction ch. III).
- La troisième phrase enfin, également source de silence, met en jeu une conjonction de coordination "et" dont l'étude n'a été que très partiellement abordée dans le présent travail.

En effet, les exemples suivants cités par GREVISSE (62, p.766 et 772) , nous ont dissuadés de rechercher dans l'immédiat des règles pouvant aider à la détermination des valeurs du nombre dans les cas des sujets multiples (voir également J.H. JAYEZ, 70, p. 147).

(1) : *Quand le prince des pasteurs et le Pontife éternel apparaîtra... (Bossuet)*

(2) : L'affection ou la haine change la justice de face. (Pascal)

Ces phrases présentent, comme on peut le voir, des schémas syntaxiques sur lesquels il est aisé de construire d'autres phrases où les accords en nombre seraient différemment établis.

PHRASE NO : 13 ELLE COMMENCE AU MOT NO: 257

DANS UN CERTAIN NOMBRE DE CAS LA FIXATION SUR LES ORGANISMES VIVANTS D' #EL#EMENTS FR#ESENTS EN PETITES QUANTIT#ES DANS L' EAU DE MER SE RAPPROCHE , PAR SES M#ECANISMES , DE CELLE QUI SE PRODUIT SUR LES CORPS INERTES : L' ESP@ECE APPORTE UNE CONTRIBUTION PASSIVE @A LA FIXATION DE L' #EL#EMENT ET LES PROPRI#ET#ES PHYSICO-CHIMIQUES DES CONTAMINANTS ET DU SUPPORT CONTAMIN#E PEUVENT PRENDRE LE PAS SUR LES PROPRI#ET#ES BIOLOGIQUES DE CE DERNIER , EN PARTICULIER SUR LA BARRI@ERE PHYSIOLOGIQUE QUE REPR#ESENTE LA R#EGULATION .

			REPR#ESENTE		S REPR#ESENTE
298 11	VS	UNE	CONTRIBUTION	19 V	
332 19			REPR#ESENTE		
332 19	SP		REPR#ESENTE	20 V	
334 20		LA	R#EGULATION		

fig. 48. (a) Mauvaise reconnaissance du sujet due à une inversion. Nous sommes en présence d'un silence partiel. Il y a également, comme nous le verrons, une mauvaise reconnaissance du complément direct (cf. ch. III, § IV. 2. 3. 2. a, fig. 57).

PHRASE NO :119 ELLE COMMENCE AU MOT NO: 3995

$00001010 DES INSTRUCTIONS PARTICULI@ERES FIXERONT LES CONDITIONS SELON LESQUELLES DES SYST@EMES M#ECANIQUES POURRONT -ETRE UTILIS#ES #EVENTUELLEMENT COMME SYST@EME DE VENTILATION D' APPOINT .

			FIXERONT		S FIXERONT	SUJET ?
3999 2	SP		FIXERONT	3 V		
4001 3		LES	CONDITIONS			

fig. 48. (b) Silence total dû à une résolution grammaticale incorrecte du mot "des" qui a reçu la catégorie grammaticale "préposition article" au lieu de "article indéfini" (cf. introduction ch. III)

PHRASE NO :138 ELLE COMMENCE AU MOT NO: 4606

L' ESCALIER CONDUISANT AUX #ETAGES ET LE VESTIBULE D' ACC@ES NE DOIVENT PAS COMPORTER DE COMMUNICATION DIRECTE AVEC LES GARAGES OU ENSEMBLES DE BOXES POUVANT R#EUNIR PLUS DE CINQ V#EHICULES AUTOMOBILES , AVEC LES MAGASINS DE VENTE ET D#EP-OTS INSTALL#ES AU REZ-DE-CHAUSS#EE OU AVEC LES D#EPENDANCES DE CES LOCAUX .

			NE DOIVENT PAS COMPORTER		S COMPORTER	SUJET ?

fig. 48. (c) Silence total dû au non-traitement de la conjonction de coordination "et" qui affecte la valeur du nombre du sujet. L'accord en nombre n'étant pas vérifié, aucun sujet potentiel n'est reconnu.

fig. 48. Bruits et silences enregistrés lors de la reconnaissance du sujet dans les formes MP.

Ainsi, à rapprocher de (2), la phrase suivante, également citée par GREVISSE (62, p. 772) :

(3) : Le bonheur ou le conseil d'autrui peuvent préserver de certaines fautes un homme très médiocre. (Fénelon, Télémaque)

2°) Reconnaissance dans les formes MIP :

• **Limitation du champ d'investigation :**

Constituent barrières :

- toute chaîne verbale d'une part ;
- tout élément libre d'une chaîne nominale d'autre part.

• **Conditions sur les éléments candidats sujets :**

Aucune condition particulière ne peut être émise quant à la nature des sujets (substantifs ou pronoms) potentiels.

Contrairement à ce qui précède, il s'agit bien cette fois d'établir davantage une liste de candidats possibles qu'une reconnaissance sûre du sujet.

Nous donnons figure 50 des exemples de relations VS obtenues automatiquement et mettant en jeu des participes présents.

Le tableau suivant donne à titre indicatif le nombre de telles relations comptabilisées sur un texte d'environ 5000 mots, ainsi que le nombre des silences enregistrés, dont la figure 51 montre les phrases qui en ont été la source.

Relations reconnues	Silences
33	2
94 %	6 %

fig. 49

3°) Reconnaissance dans les formes MII :

Ces formes étant rarement utilisées avec un sujet exprimé, nous avons, du moins dans l'immédiat, délaissé la recherche dans ce cas.

PHRASE NO : 32 ELLE COMMENCE AU MOT NO: 812

LA CONVERSION DE LA CHALEUR EN #ELECTRICIT#E SE FAIT PAR UN SYST∂EME DYNAMIQUE COMPOS#E DE DEUX TURBINES ∂A VAPEUR DE LIQUIDE ORGANIQUE , TRAVAILLANT EN CASCADE .

TRAVAILLANT EN CASCADE S TRAVAILLANT

823	4	VS		UN	SYST∂EME	8 D
828	5	VS	DE	DEUX	TURBINES	8 D
830	6	VS	∂A		VAPEUR	8 D
832	7	VS	DE		LIQUIDE	8 D
835	8				TRAVAILLANT	

PHRASE NO : 52 ELLE COMMENCE AU MOT NO: 6693

$00001510 NONOBSTANT LES DISPOSITIONS DE L' ARR-ET#E DU 14 D#ECEMBRE 1961 PR#ECIT#E , LA DEMANDE DE PERMIS DE CONSTRUIRE EST ACCOMPAGN#EE DES PI∂ECES SUIVANTES : 1.- PLAN DE SITUATION ∂A L' #ECHELLE DU CADASTRE , COMPORTANT L' IMPLANTATION DE LA CONSTRUCTION PROJET#EE ; 2.- CROQUIS COT#E DE LA CONSTRUCTION PROJET#EE , ANNOT#E DE LA NATURE DES MAT#ERIAUX PR#EVUS .

COMPORTANT S COMPORTANT

6719	6	VS			PLAN	10 D
6721	7	VS	DE		SITUATION	10 D
6724	8	VS	∂A	L'	#ECHELLE	10 D
6726	9	VS	OU		CADASTRE	10 D
6728	10				COMPORTANT	

PHRASE NO : 60 ELLE COMMENCE AU MOT NO: 7039

LE R#ESULTAT DE CE CALCUL EST AFFECT#E DES POURCENTAGES FIGURANT AU TABLEAU CI-APR∂ES .

FIGURANT S FIGURANT

7047	3	VS	DES		POURCENTAGES	4 V
7048	4				FIGURANT	

PHRASE NO : 61 ELLE COMMENCE AU MOT NO: 1664

TOUTE VACANCE SURVENANT EN COURS DE MANDAT PEUT ~ETRE IMM#EDIATEMENT COMBL#EE PAR UNE NOMINATION PRONONC#EE POUR LA DUR#EE DU MANDAT RESTANT ∂A COURIR .

SURVENANT S SURVENANT

1665	1	VS		TOUTE	VACANCE	2 V
1666	2				SURVENANT	

PHRASE NO : 84 ELLE COMMENCE AU MOT NO: 7618

LES PR-ETS B#EN#EFICIANT DE LA GARANTIE DE L' &ETAT DANS LES LIMITES ET CONDITIONS FIX#EES PAR LE D#ECRET N. 63-1324 DU 24 D#ECEMBRE 1963 MODIFI#E , SERONT R#EALIS#ES : S' ILS SONT ACCORD#ES AU TITRE DU PARAGRAPHE 1 DE L' ARTICLE 25 DU D#ECRET SUSVIS#E : SOUS LA FORME SOIT DE PR-ETS ∂A MOYEN OU ∂A LONG TERME ∂A CONSENTIR PAR LE &CR#EDIT FONCIER DE &FRANCE , SOIT D' OUVERTURES DE CR#EDIT ∂A CONSENTIR PAR LE &COMPTOIR DES ENTREPRENEURS AGISSANT SEUL OU AVEC LE CONCOURS DU &CR#EDIT FONCIER DE &FRANCE .

B#EN#EFICIANT S B#EN#EFICIANT

7619	1	VS		LES	PR-ETS	2 V
7620	2				B#EN#EFICIANT	

AGISSANT SEUL S AGISSANT

7692	17	VS		LE	&COMPTOIR	19 D
7694	18	VS	DES		ENTREPRENEURS	19 D
7695	19				AGISSANT	

Fig. 50. Reconnaissance des sujets potentiels dans les formes MIP.

```
PHRASE NO :  8    ELLE COMMENCE AU MOT NO:  4952

@A D#EFAUT , IL DOIT -ETRE PR#EVU DES DISPOSITIFS PARTICULIERS DE CONSTRUCTION ( COURSIVES @A L' AIR
LIERE @A CERTAINS NIVEAUX NOTAMMENT ) PERMETTANT D' ASSURER L' #EVACUATION DES PERSONNES .

PERMETTANT D' ASSURER                                                S ASSURER                SUJET

4976  6  SP                          ASSURER                  7 V
4978  7                 L'           #EVACUATION
```

fig. 51. (a). Premier cas de silence. Il est total. Il y a lieu ici de reconnaître la structure parenthésée afin de procéder à sa suppression virtuelle. Elle joue en effet actuellement un rôle de barrière.

```
PHRASE NO :122    ELLE COMMENCE AU MOT NO:  4112

$00001050 LES INSTALLATIONS FIXES DE CHAUFFAGE NE DOIVENT OCCASIONNER @A L' INT#ERIEUR DES LOCAUX
AUCUN D#EGAGEMENT DE FUM#EE OU DE GAZ POUVANT COMPROMETTRE LA SANT#E DES HABITANTS .

POUVANT COMPROMETTRE                                                 S COMPROMETTRE

4129  6  VS   DE                     GAZ                      7 V
4131  7                              COMPROMETTRE
```

fig. 51. (b). Deuxième cas de silence. Il est dû, comme précédemment (cf. fig. 48. (c).) au non-traitement des conjonctions de coordination lorsque celles-ci portent sur des chaînes nominales.
Ce silence est partiel. Nous verrons que ce sont ces cas là qui pourront engendrer des erreurs dans l'étape future du système, c'est à dire lors de l'apprentissage sémantique.

fig. 51. Silences enregistrés lors de la reconnaissance du sujet dans les formes MIP.

Indiquons néanmoins que nous avons relevé trois cas de silence (cf. fig.52) sur environ 40 formes infinitives répertoriées.

b) Dans les formes passives :

Ici le sujet est placé après le verbe. Sa reconnaissance appellera néanmoins, comme précédemment, une limitation de la portée de recherche d'une part, et un ensemble de conditions à respecter portant sur les candidats sujets d'autre part.

- **Limitation du champ d'investigation :**

Les barrières qui ont été retenues pour le moment sont constituées des éléments suivants :

- les chaînes verbales à la forme active ;
- les conjonctions de subordination ou de coordination ;
- certaines ponctuations (, ; :) ;
- les éléments sujets potentiels des formes actives de type MP enfin.

- **Conditions portant sur les candidats sujets :**

Tout comme dans les chaînes nominales (cf. ch. III, § IV. 2. 1. 2. d) nous n'avons retenu que les éléments introduits par la préposition "par", en ajoutant cependant qu'ils doivent être de surcroît libres.

Rappelons en effet que pour retenir la préposition "de" en tant qu'introducteur possible du sujet dans les formes passives, il nous aurait fallu dresser une liste des verbes pouvant avoir leur sujet introduit par "de". Or nous ne disposions pas de cette liste. Notre démarche fut donc de ne retenir que la préposition "par" qui, contrairement à la préposition "de", introduit presque toujours le sujet dans les structures considérées.

Nous n'avons enregistré ainsi que deux cas de bruits totaux (cf. fig. 53) dont nous donnons figure 55 les phrases qui en ont été le siège.

La figure 54 offre quelques exemples de reconnaissance correctement effectuée.

A titre indicatif, nous consignons dans le tableau suivant les résultats chiffrés que nous avons comptabilisés sur un texte d'environ 5000 mots.

```
PHRASE NO :  4   ELLE COMMENCE AU MOT NO:  4812

CELLES DESTIN#EES AUX COLONNES MONTANTES DE GAZ DOIVENT ¬ÊTRE OUVERTES @A LEUR PARTIE INF#ERIEURE ET
COMMUNIQUER AVEC L' ATMOSPH@ERE @A LEUR PARTIE SUP#ERIEURE .

ET COMMUNIQUER                                          S COMMUNIQUER                SUJET ?
```

fig. 52. (a). Silence dû au non-traitement des conjonctions de coordination portant sur des chaînes verbales.

```
PHRASE NO : 22   ELLE COMMENCE AU MOT NO:  5290

$00001340 SOUS R#ESERVE DES DISPOSITIONS PR#EVUES AUX ARTICLES CI-APR@ES , LES ORGANISMES D'
HABITATIONS @A LOYER MOD#ER#E NE PEUVENT , SAUF D#EROGATION ACCORD#EE PAR LE MINISTRE DE LA
CONSTRUCTION , APR@ES AVIS DU CONSEIL SUP#ERIEUR DES HABITATION @A LOYER MOD#ER#E , IMPOSER AUX
B#EN#EFICIAIRES DE LEUR ACTIVIT#E DES CHARGES FINANCI@ERES SUP#ERIEURES @A CELLES QU' ILS SUPPORTENT
ANNUELLEMENT POUR L' AMORTISSEMENT ET L' INT#ER¬ET DES PR¬ETS QU' ILS ONT CONTRACT#ES EN VUE D'
OP#ERATIONS D' ACCESSION @A LA PROPRI#ET#E .

IMPOSER                                                 S IMPOSER                    SUJET ?
```

fig. 52. (b). Silence dû à une reconnaissance incomplète de la chaîne verbale (cf. ch. III, § III. 2.).

```
PHRASE NO :132   ELLE COMMENCE AU MOT NO:  4381

$00001140 CHAQUE LOGEMENT DOIT COMPORTER UN #EQUIPEMENT PERMETTANT DE LE CHAUFFER SUFFISAMMENT OU
¬ETRE SUSCEPTIBLE DE RECEVOIR CET #EQUIPEMENT .

OU ¬ETRE SUSCEPTIBLE DE RECEVOIR                        S RECEVOIR                   SUJET ?
```

fig. 52. (c). Silence dû également au non-traitement des conjonctions de coordination.

fig. 52. Silences provoqués par la non-recherche des sujets potentiels pour les formes MII.

Reconnaissance correcte	Bruit Total
18	2
90 %	10 %

fig. 53. Sujet introduit par "par"

Indiquons enfin que retenir la préposition "de" nous aurait alors conduit au résultat consigné dans le tableau suivant :

Reconnaissance correcte	Bruit Total
6	3
67 %	33 %

fig. 56. Sujet introduit par "de".

Or rappelons que nous ne maintenons une option que si le taux du bruit total auquel elle est susceptible de conduire n'excède pas le seuil fixé arbitrairement à 10 % pour la relation considérée.

Nous retiendrons donc finalement que nous avons reconnu ici, non pas le sujet potentiel, mais le sujet d'une façon sûre (cf. ch. V, § I. 3.).

IV. 2. 3. 2. Reconnaissance des relations VD potentielles :

a) Dans les formes actives :

Les options que nous avons retenues ici sont fort simplificatrices, mais, comme nous le verrons, elles n'introduiront que très peu de bruit ou silence.

```
PHRASE NO :  4   ELLE COMMENCE AU MOT NO:    82

CE RADIONUCL#EIDE EST SUSCEPTIBLE DE SE FIXER SUR LES PARTICULES EN SUSPENSION QUI SONT PI#EG#EES
PAR LES ESP∂ECES FILTREUSES TELLES QUE LES MOULES ( &MYTILUS &EDULIS &L. ) .

SONT PI#EG#EES                                                          P PI#EG#EES

 96  4                          PI#EG#EES
 99  5  VS          LES         ESP∂ECES                  4 V
-------------------------------------------------------------------------------------------------

PHRASE NO : 47   ELLE COMMENCE AU MOT NO:  1291

$00000020 L' #ECHANGE PR#EVU ∂A L' ARTICLE 219 DU CODE DE L' URBANISME ET DE L' HABITATION POURRA
¬ETRE IMPOS#E PAR L' ORGANISME D' HABITATIONS ∂A LOYER MOD#ER#E , PROPRI#ETAIRE , ∂A CEUX DE SES
LOCATAIRES OU OCCUPANTS QUI SE TROUVENT DANS LES CONDITIONS D' OCCUPATION INSUFFISANTE D#EFINIES ∂A
L' ARTICLE 2 CI-APR∂ES EN VUE D' OCCUPER UN LOGEMENT SITU#E SUR LE TERRITOIRE DE LA M¬EME COMMUNE OU
D' UNE COMMUNE LIMITROPHE .

POURRA ¬ETRE IMPOS#E                                                    P IMPOS#E

1310  6                          IMPOS#E
1313  7  VS          L'          ORGANISME                 6 V
-------------------------------------------------------------------------------------------------

PHRASE NO : 64   ELLE COMMENCE AU MOT NO:  7211

TOUTEFOIS , DES D#EROGATIONS AUX DISPOSITIONS DE L' ARTICLE 2 CI-DESSUS POURRONT ¬ETRE ACCORD#EES
PAR LE MINISTRE DE LA CONSTRUCTION POUR LES DOSSIERS DONT LA DEMANDE DE PERMIS DE CONSTRUIRE AURA
#ET#E D#EPOS#EE ANT#ERIEUREMENT ∂A LA PUBLICATION DU PR#ESENT ARR¬ET#E .

POURRONT ¬ETRE ACCORD#EES                                               P ACCORD#EES

7224  4                          ACCORD#EES
7227  5  VS          LE          MINISTRE                  4 V
7233  7  VC  POUR   LES         DOSSIERS                  4 V
-------------------------------------------------------------------------------------------------
```

fig. 54. Reconnaissance des relations VS dans les formes passives.

```
PHRASE NO : 39   ELLE COMMENCE AU MOT NO:   946

IL RESSORT DE CETTE #ETUDE QUE LES COMPOS#ES MAGN#ESIENS ONT #ET#E PEU #ETUDI#ES PAR SPECTROSCOPIE
INFRAROUGE ET QUE L' ON POSSDEDE PEU DE R#ESULTATS EN CE QUI CONCERNE CES COMPLEXES .

     ONT #ET#E PEU #ETUDI#ES                                          P #ETUDI#ES

953  2                LES       COMPOS#ES
953  3  SP                      #ETUDI#ES                 2 V

953  3                          #ETUDI#ES
960  4  VS                      SPECTROSCOPIE             3 V
```

fig. 55. (a). Bruit total. Notre algorithme ne peut détecter cette erreur puisque **"étudier"** admet pour introducteur du sujet la préposition **"par"** . Cependant, nous pourrions faire intervenir l'absence d'article pour rejeter **"par spectroscopie"** comme expression du complément d'agent.

```
PHRASE NO :121   ELLE COMMENCE AU MOT NO:  4061

$00001020 LES PRESCRIPTIONS CI-DESSUS SONT FAITES SOUS R#ESERVE DES R@EGLES PARTICULI@ERES QUI
POURRONT ¬ETRE #EDICT#EES EN FONCTION DES RISQUES D' INTOXICATION PAR LE GAZ OU ,
D'UNE MANI@ERE G#EN#ERALE , PAR LES APPAREILS DE COMBUSTION SITU#ES DANS LES LOCAUX D' HABITATION ET
POUR ASSURER @A CES APPAREILS LES QUANTIT#ES D' AIR N#ECESSAIRES @A LEUR BON FONCTIONNEMENT .

4073  3                          #EDICT#EES

4080  6  VS            LE        GAZ                       3 V
```

fig. 55. (b). Bruit total. L'erreur ici est due à l'analyse incomplète de la chaîne nominale pour laquelle nous avons admis que **"par"** constituait une barrière. L'unité lexicale **"gaz"** est donc restée libre. D'où le bruit enregistré.

fig. 55. Bruits totaux enregistrés lors de la reconnaissance des relations VS dans les formes passives.

Ajoutons qu'elles sont désormais communes à toutes les formes actives que nous avons distinguées.

• **Limitation du champ d'investigation :**

Constituent barrières les éléments suivants :

- les chaînes verbales à la forme active ;
- les conjonctions de subordination ou de coordination ;
- les ponctuations (, ; :) ;
- les pronoms sujets enfin.

Notons que ces barrières sont semblables à celles qui ont été mises en oeuvre lors de la reconnaissance du sujet dans les formes passives.

• **Conditions sur les candidats compléments directs :**

Ils sont, comme les candidats sujets des formes actives :

- libres, d'une part, et
- ne possèdent pas de relation satellite prépositive, d'autre part.

Ces deux conditions étant satisfaites, c'est de plus l'élément le plus proche de la chaîne verbale qui est pris en compte.

Ces options permettent de reconnaître le complément direct dans un grand nombre de cas; ainsi que le montre le tableau suivant où nous avons consigné les résultats chiffrés que nous avons relevés sur un texte d'environ 5000 mots.

Relations reconnues	Silence Total	Bruit Total
92	8	1
91 %	8 %	1 %

fig. 57. Relations VD dans les formes actives.

Les figures 58 et 59 donnent respectivement des exemples de reconnaissance et de non-reconnaissance (silence total), la phrase donnant lieu au seul bruit total enregistré étant donné figure 48 (a).

PHRASE NO : 16 ELLE COMMENCE AU MOT NO: 410

$ LORS DU TRAITEMENT DES COMBUSTIBLES IRRADI#ES , L' USINE DE &LA &HAGUE A REJET#E EN MER DES EFFLUENTS RADIOACTIFS CONSTITU#ES DE 137&CS ET DE 106&RU .

******* CHAINES VERBALES RECONNUES

A REJET#E S REJET#E

418	3	VS	L'	USINE	5	V
422	5			REJET#E		
422	5	SP		REJET#E	7	V
426	7		DES	EFFLUENTS		

PHRASE NO :120 ELLE COMMENCE AU MOT NO: 4018

CES INSTRUCTIONS PR#ECISERONT LES R@EGLES DE S#ECURIT#E , EN VUE NOTAMMENT D' #EVITER QUE L' ACTION DE CES SYST@EMES NE VIENNE SUPPRIMER LE TIRAGE DES APPAREILS DE COMBUSTION SITU#ES DANS LE LOGEMENT ET RACCORD#ES @A DES CONDUITS DE FUM#EE OU @A DES VENTOUSES .

******* CHAINES VERBALES RECONNUES

PR#ECISERONT S PR#ECISERONT

4019	1	VS	CES	INSTRUCTIONS	2	V
4020	2			PR#ECISERONT		
4020	2	SP		PR#ECISERONT	3	V
4022	3		LES	R@EGLES		

EN VUE NOTAMMENT D' #EVITER S #EVITER SUJET ? COMPLEMENT

NE VIENNE SUPPRIMER S SUPPRIMER

4032	4	VS	L'	ACTION	6	V
4038	6			SUPPRIMER		
4038	6	SP		SUPPRIMER	7	V
4040	7		LE	TIRAGE		

PHRASE NO :122 ELLE COMMENCE AU MOT NO: 4112

$00001050 LES INSTALLATIONS FIXES DE CHAUFFAGE NE DOIVENT OCCASIONNER @A L' INT#ERIEUR DES LOCAUX AUCUN D#EGAGEMENT DE FUM#EE OU DE GAZ POUVANT COMPROMETTRE LA SANT#E DES HABITANTS .

******* CHAINES VERBALES RECONNUES

NE DOIVENT OCCASIONNER S OCCASIONNER

4114	1	VS	LES	INSTALLATIONS	3	V
4120	3			OCCASIONNER		
4120	3	SP		OCCASIONNER	5	V
4124	5		AUCUN	D#EGAGEMENT		

POUVANT COMPROMETTRE S COMPROMETTRE

4129	6	VS	DE	GAZ	7	V
4131	7			COMPROMETTRE		
4131	7	SP		COMPROMETTRE	8	V
4133	8		LA	SANT#E		

fig. 58. Reconnaissance des relations VD (notées SP par homogénéîté avec les mêmes relations établies dans les chaînes nominales) dans les formes actives.

```
      PHRASE NO : 15   ELLE COMMENCE AU MOT NO:  3400

      CES CONSTRUCTIONS QUI FONT APPEL @A DES PRINCIPES NON TRADITIONNELS FR#ESENTENT SOIT DANS L'
      AGENCEMENT INT#ERIEUR DES PLANS , SOIT DANS LES DIMENSIONS DE BASE DES LOCAUX , SOIT DANS L'
      #EQUIPEMENT DES PIECES PRINCIPALES OU SECONDAIRES , UN CARACT@ERE NOVATEUR PAR RAPPORT AUX USAGES
      ET @A LA R#EGLEMENTATION EN VIGUEUR .

****** CHAINES VERBALES RECONNUES

      FONT APPEL                                                    S FONT APPEL

3401  1  VS            CES       CONSTRUCTIONS          2 V
3403  2                         FONT APPEL
-------------------------------------------------------------------------------------------------
      FR#ESENTENT                                                   S FR#ESENTENT

3401  1  VS            CES       CONSTRUCTIONS          3 V
3409  3                         FR#ESENTENT                                              COMPLEMEN
-------------------------------------------------------------------------------------------------

      PHRASE NO : 22   ELLE COMMENCE AU MOT NO:  3629

      UN ARR-ET#E DU PR#EFET D#EFINIRA , SELON LES AGGLOM#ERATIONS , L' #ELOIGNEMENT MINIMUM @A
      PRENDRE EN CONSID#ERATION POUR L' ATTRIBUTION DE CETTE MAJORATION .

****** CHAINES VERBALES RECONNUES

      D#EFINIRA                                                     S D#EFINIRA

3630  1  VS            UN        ARR-ET#E               3 V
3633  3                         D#EFINIRA                                                COMPLEMEN
-------------------------------------------------------------------------------------------------

      PHRASE NO : 92   ELLE COMMENCE AU MOT NO:  3054

      $00000490 LE PR#ESIDENT DE LA CONF#ERENCE PEUT ENTENDRE , POUR LES AFFAIRES QUI LES CONCERNENT ,
      TOUTES AUTORIT#ES OU PERSONNES COMP#ETENTES POUR #EMETTRE UN AVIS SUR CES AFFAIRES .

****** CHAINES VERBALES RECONNUES

      PEUT ENTENDRE                                                 S ENTENDRE

3056  1  VS            LE        PR#ESIDENT             3 V
3061  3                         ENTENDRE                                                 COMPLEMEN
-------------------------------------------------------------------------------------------------
      CONCERNENT                                                    S CONCERNENT

3065  3  VS   POUR     LES       AFFAIRES               4 V
3068  4                         CONCERNENT                                               COMPLEMEN
-------------------------------------------------------------------------------------------------
```

fig. 59. Silences totaux enregistrés lors de la reconnaissance des relations VD dans les formes actives

Remarques : Tous les silences que nous avons relevés sont analogues à ceux ici exposés. Tous en effet sont dûs à des insertions entre le verbe et son complément d'une ou plusieurs propositions de structures diverses. Ajoutons que celles ci demandent, pour être reconnues, une analyse fondée sur une segmentation non pas en chaînes nominales et verbales, mais en propositions (cf. ch. III, § III). La suppression fictive de ces propositions dès lors possible, permettra, en reprenant les présents traitements, la détermination des compléments directs dans des structures syntaxiques où ceux-ci seraient relativement éloignés du verbe auquel ils se rapportent. (voir également fig. 63, silence dans VI_p).

b) Dans les formes passives :

Ce sont en fait les options déjà formulées en § IV. 2. 3. 1., pour la reconnaissance du sujet dans les formes actives, qui sont ici reprises. Les procédures de reconnaissance sont donc les mêmes avec en plus, cependant, la prise en compte du genre.

Les remarques que nous avons faites en § IV. 2. 3. 1., relatives à la reconnaissance du sujet dans les formes MP, peuvent donc être reprises ici. Nous avons en particulier relevé des sources de bruit ou de silence analogues.

Nous donnons dans le tableau suivant les performances de l'algorithme mesurées sur un texte d'environ 5000 mots.

Nature du complément direct												Performances globales		
Substantifs				Pronoms				Pr. rel. qui						
Rec. cor.	Sil. total		Bruit	Rec. cor.	Sil total		Br.	Rec. cor.	Sil total		Br.	Rec. cor.	Sil. total	Br.
44	3		3	9	0		0	6	0		0	59	3	3
88 %	6 %		6 %	100 %	-		-	100 %	-		-	91 %	4.5 %	4.5 %

fig. 60 - Reconnaissance des relations VD dans les formes passives MP.

La figure 61 offre des exemples de reconnaissance de relations VD dans les formes passives MP. Nous n'avons en effet relevé aucune forme passive introduite par un paricipe présent ou un infinitif.

IV. 2. 3. 3. Reconnaissance des relations VI_p potentielles :

Bien que les options que nous allons donner ici puissent être communes, à quelques différences près, aux formes actives ou passives, elles n'ont été en fait mises à l'épreuve que dans le cas des formes passives.

```
        PHRASE NO : 59   ELLE COMMENCE AU MOT NO: 1633

        $00000110 LES MEMBRES DU CONSEIL SUP#ERIEUR DU MINIST∂ERE DE LA CONSTRUCTION SONT NOMM#ES POUR
        DEUX ANS PAR ARR-ET#E DU MINISTRE DE LA CONSTRUCTION .

        SONT NOMM#ES                                                    P NOMM#ES

  1635  1                LES       MEMBRES
  1645  5  SP                      NOMM#ES                  1 V
```

```
        PHRASE NO :110   ELLE COMMENCE AU MOT NO:  8881

        LE TARIF DES DROITS D' INSTRUCTION SE RAPPORTANT AUX PR-ETS VIS#ES ∂A L' ARTICLE 3 DES ARR-ET#ES
        SUSVIS#ES DES 28 D#ECEMBRE 1963 , 20 MAI 1965 ET 13 JUILLET 1965 EST FIX#E FORFAITAIREMENT ∂A 50 &F
        PAR LOGEMENT .

        EST FIX#E FORFAITAIREMENT                                       P FIX#E

  8882  1                LE        TARIF
  8912  7  SP                      FIX#E                    1 V


        PHRASE NO :112   ELLE COMMENCE AU MOT NO:  3698

        $00000930 SI LES M#EMOIRES NE SONT PAS PRODUITS DANS UN D#ELAI DE DEUX ANS ∂A COMPTER DE LA
        NOTIFICATION DE LA D#ECISION PROVISOIRE D' OCTROI DE PRIMES , CELLE-CI EST ANNUL#EE .

        NE SONT PAS PRODUITS                                            P PRODUITS

  3701  1                LES       M#EMOIRES
  3705  2  SP                      PRODUITS                 1 V                                  C. D'AGENT ?

-------------------------------------------------------------------------------------------------------------
        EST ANNUL#EE                                                    P ANNUL#EE
COMPL DE TYPE PRONOMINAL :  CELLE-CI
```

```
        PHRASE NO :117   ELLE COMMENCE AU MOT NO:  3832

        $00001000 LES EMPLACEMENTS DES OUVERTURES , TANT EN CE QUI CONCERNE LES CUISINES QUE LES PI∂ECES
        PRINCIPALES D' HABITATION , DOIVENT -ETRE #ETUDI#ES POUR QUE CES OUVERTURES B#EN#EFICIENT AU MIEUX
        DES DIFF#ERENCES DE PRESSION ENTRE L' INT#ERIEUR ET L' EXT#ERIEUR PROVENANT TANT DE LA FORCE
        DYNAMIQUE DES COURANTS EXT#ERIEURS QUE DES GRADIENTS DE TEMP#ERATURE .

        DOIVENT -ETRE #ETUDI#ES                                         P #ETUDI#ES

  3884  1                LES       EMPLACEMENTS
  3901  6  SP                      #ETUDI#ES                1 V                                  C. D'AGENT ?
-------------------------------------------------------------------------------------------------------------

        PHRASE NO :127   ELLE COMMENCE AU MOT NO:  4267

        ILS DOIVENT ALORS -ETRE SOLIDAIREMENT FIX#ES ∂A LEUR SUPPORT .

        DOIVENT ALORS -ETRE SOLIDAIREMENT FIX#ES                        P FIX#ES
COMPL DE TYPE PRONOMINAL :  ILS

                                                                                                  C. D'AGENT ?
  4272  1                         FIX#ES
  4275  2  VC   ∂A        LEUR    SUPPORT                  1 V
-------------------------------------------------------------------------------------------------------------
```

fig. 61. Reconnaissance de relations VD (notées SP) dans les formes passives.

La recherche du complément indirect potentiel est dans ce cas analogue à la recherche du sujet dans les formes passives.

En particulier la limitation du champ d'investigation met en oeuvre les mêmes barrières (cf. ch. III, § IV. 2. 3. 1. b).

Les conditions portant sur les candidats compléments indirects sont par contre :

- qu'ils soient libres d'une part ;
- et qu'ils possèdent une relation satellite prépositive d'autre part ; la préposition mise en jeu étant cependant différente de "par".

C'est enfin toujours l'élément répondant à ces conditions le plus proche géographiquement, qui est pris en compte.

Le tableau suivant donne les performances de la procédure utilisée, comptabilisée sur un texte d'environ 5000 mots.

Reconn. correcte	Silence Total	Bruit Total
43	3	0
93,5 %	6,5 %	-

fig. 62. Relations VI_p dans les formes passives.

Remarque :

Nous n'avons pas comptabilisé comme étant des bruits les cas (au nombre de 5) de relations VI_p mettant en jeu en fait un sujet introduit par "de". Car l'erreur porte alors sur la nature du lien et non sur son existence. De plus, comme nous le verrons, la simulation des règles transformationnelles rendra possible la rectification de telles anomalies.

La figure 64 offre des exemples de reconnaissance de relations VI_p dans les formes passives.

La figure 63 offre par contre un exemple de silence total. Les barrières mises en oeuvre lors de la reconnaissance des relations VI_p étant semblables à celles mises en oeuvre lors de la reconnaissance des relations VD dans les formes actives, les remarques que nous avons faites alors peuvent donc être reprises ici.

```
PHRASE NO :  1   ELLE COMMENCE AU MOT NO:  1290

$00000020 L' #ECHANGE PR#EVU ∂A L' ARTICLE 219 DU CODE DE L' URBANISME ET DE L' HABITATION POURRA
-ETRE IMPOS#E PAR L' ORGANISME D' HABITATIONS ∂A LOYER MOD#ER#E , PROPRI#ETAIRE , ∂A CEUX DE SES
LOCATAIRES OU OCCUPANTS QUI SE TROUVENT DANS LES CONDITIONS D' OCCUPATION INSUFFISANTE D#EFINIES ∂A
L' ARTICLE 2 CI-APR∂ES EN VUE D' OCCUPER UN LOGEMENT SITU#E SUR LE TERRITOIRE DE LA M-EME COMMUNE OU
D' UNE COMMUNE LIMITROPHE .

1290  1                L'        #ECHANGE
1309  6  SP                      IMPOS#E                    1 V

1309  6                          IMPOS#E
1312  7  VS            L'        ORGANISME                  6 V
-------------------------------------------------------------------------------------------------
```

fig. 63. Silence total enregistré lors de la reconnaissance des relations VI_p dans les formes passives. Seules ont été reconnues les relations VD (notées SP) et VS de la forme passive détectée.

IV. 2. 4. Analyse de l'anaphorique "qui":

Il sagit ici de reconnaître l'antécédent, ou, en reprenant la terminologie proposée par L. TESNIERE, la source sémantique qui communique sa valeur à l'élément anaphorique considéré.

Nous ne saurions cependant prétendre ici à l'analyse de l'anaphore dans a totalité, dont le traitement déborde souvent le cadre de la phrase (cf. ch. I, § IV.). Aussi nous limiterons-nous à la seule reconnaissance des antécédents potentiels du pronom relatif "qui".

Notre but est en effet d'illustrer simplement la méthodologie que nous proposons.

Notons toutefois que notre choix, qui s'est fixé sur le pronom relatif "qui", a été guidé par deux considérations :

- la première est relative au cadre que nous nous sommes fixé pour l'analyse, à savoir la phrase. Il fallait donc que l'antécédent figurât dans la même phrase que l'anaphorique qui le reprend ;
- la deuxième est relative à la fréquence d'apparition de l'anaphorique. Il fallait en effet qu'il soit relativement répandu.

IV. 2. 4. 1. Reconnaissance de l'antécédent :

"Toute anaphore suppose deux connexions sémantiques, 1° celle qui double la connexion structurale, et 2° la connexion sémantique supplémentaire qui constitue l'anaphore". (L. TESNIERE, 108, p. 85) .

```
PHRASE NO :  1   ELLE COMMENCE AU MOT NO:  2904

$00000330 LES EXEMPTIONS PR#EVUES PAR LE PR#ESENT ARR-ET#E NE SONT PAS APPLICABLES AUX TRAVAUX
CONCERNANT LES CONSTRUCTIONS FRAPP#EES D' ALIGNEMENT ET CELLES SITU#EES DANS LE P#ERIM@ETRE DE
PROTECTION DES MONUMENTS HISTORIQUES ET DES SITES CLASS#ES .

******* CHAINES VERBALES RECONNUES

        NE SONT PAS APPLICABLES                                    P APPLICABLES

 2906  1                LES       EXEMPTIONS
 2915  3  SP                      APPLICABLES              1 V                        C. D'AGENT

 2915  3                          APPLICABLES
 2917  4  VC   AUX               TRAVAUX                  3 V
-------------------------------------------------------------------------------------------------
        CONCERNANT                                                 S CONCERNANT

 2917  3  VS   AUX               TRAVAUX                  4 V
 2918  4                        CONCERNANT

 2918  4  SP                    CONCERNANT               5 V
 2920  5                LES     CONSTRUCTIONS
-------------------------------------------------------------------------------------------------


        PHRASE NO :  3   ELLE COMMENCE AU MOT NO:  9473

        ENFIN , CERTAINES ACTIVIT#ES , NORMALEMENT SITU#EES HORS DU CHAMP D' APPLICATION DE LA &T.&V.&A. ,
        PEUVENT -ETRE ASSUJETTIES @A CETTE TAXE SUR LA DEMANDE EXPRESSE DES ENTREPRISES INT#ERESS#EES : C'
        EST LE R#EGIME DE L' OPTION .

******* CHAINES VERBALES RECONNUES

        PEUVENT -ETRE ASSUJETTIES                                  P ASSUJETTIES

 9476  1                CERTAINES ACTIVIT#ES
 9490  5  SP                      ASSUJETTIES              1 V                        C. D'AGENT

 9490  5  SP                      ASSUJETTIES
 9493  6  VC   @A         CETTE   TAXE                     5 V
-------------------------------------------------------------------------------------------------
        EST                                                        0 EST                SUJET ?
-------------------------------------------------------------------------------------------------


        PHRASE NO :111   ELLE COMMENCE AU MOT NO:  3683

        $00000680 EN CAS DE FAUSSE D#ECLARATION , LE DEMANDEUR EST RAY#E DE LA LISTE DE CLASSEMENT .

.................................................................................................

******* CHAINES VERBALES RECONNUES

        EST RAY#E                                                  P RAY#E

 3689  2                LE        DEMANDEUR
 3691  3  SP                      RAY#E                    2 V                        C. D'AGENT

 3691  3                          RAY#E
 3694  4  VC   DE         LA      LISTE                    3 V
-------------------------------------------------------------------------------------------------
```

fig. 64. Reconnaissance des relations VI (notées VC) dans les formes passives.

Le premier type de connexion correspond aux liens déjà établies lors des étapes précédentes, c'est à dire lors de la reconnaissance du sujet dans les formes actives MP ou du complément direct dans les formes passives. Remarquons que ces liens matérialisaient alors des relations satellites de type hétérosyntagmatique (cf. ch. III, § IV. 1.).

Il nous reste donc à établir les secondes connexions.

La détermination de ces connexions que nous appellerons liens anaphoriques sera cependant accompagnée d'une opération de substitution : nous remplacerons en effet d'emblée l'anaphorique par la liste des antécédents potentiels qui lui sont associés en ce point.

IV. 2. 4. 2. Les options :

Nous avons là encore adopté des options fort simplificatrices. Mais, comme pour toutes les options que nous avons jusqu'ici considérées, elles n'introduisent que peu de silence ou bruit notamment.

- **Limitation du champ d'investigation :**

Constituent barrières :

- toute chaîne verbale ;
- tout élément libre d'un chaîne nominale ;
- ou enfin, toute conjonction de coordination.

- **Conditions sur les antécédents candidats :**

- lorsqu'ils sont substantifs, ne pas être déterminés (GREVISSE, 62, p. 403 et 465) . Le substantif candidat doit donc posséder une relation satellite de détermination (cf. ch. III, § IV. 1. et IV. 2. 1. 1., fig. 25).
- avoir les mêmes valeurs de genre, nombre et personne que celles dont l'anaphorique est pourvu et qu'il a lui-même hérité de la chaîne verbale à laquelle il est attaché.

IV. 2. 4. 3. Résultats de l'expérimentation :

A titre indicatif, nous trouvons dans le tableau suivant les performances de la procédure utilisée mesurées sur un texte d'environ 6000 mots.

Fonction des antécédents						Performances globales		
Sujet			Complément direct					
Recon. cor.	Sil. total	Bruit	Recon. cor.	Sil. total	Bruit	Recon. cor.	Sil. total	Bruit
23	2	-	7	1	-	30	3	-
92 %	8 %	-	87 %	13 %	-	91 %	9 %	-

fig. 65 - Reconnaissance des antécédents potentiels.

Les figures suivantes donnent respectivement :

- figures 66 (a) et 66 (b) : des exemples de reconnaissance de liens anaphoriques potentiels, accompagnée de substitutions donnant lieu à des relations VS pour des formes actives.
- figure 66 (c) : des exemples de reconnaissance de liens anaphoriques donnant lieu cette fois à des relations VD pour des formes passives.
- figure 67 (a) : les phrases qui ont donné lieu à des silences totaux.

Remarques :

La figure 67 (b) offre un cas de silence que nous n'avons pas comptabilisé et qui est dû à une erreur dans l'orthographe du mot "SINISTRE". Celui-ci aurait dû en effet, lors de la saisie du texte, être accordé en nombre avec le mot auquel il se rapporte, à savoir "LOCATAIRES".

C'est là un autre exemple, différent (cf. introduction ch. III, § IV. 2. 3. 1. a., fig. 48(b)), où les informations construites lors des présentes étapes (absence d'un complément direct dans le cas présent) peuvent aider à la détection de certaines erreurs. Ces mêmes informations peuvent alors déclencher des retours arrières et permettre la correction des erreurs détectées, ici d'ordre morphologique.

PHRASE NO : 36 ELLE COMMENCE AU MOT NO: 877

$ CETTE BIBLIOGRAPHIE CONCERNE LES COMPOS#ES M#ETALLIQUES QUI PR#ESENTENT DES LIAISONS M#ETAL-AZOTE .

******* CHAINES VERBALES RECONNUES

CONCERNE						S CONCERNE	
879	1	VS	CETTE	BIBLIOGRAPHIE	2 V		
880	2			CONCERNE			
880	2	SP		CONCERNE	3 V		
882	3		LES	COMPOS#ES			
PR#ESENTENT						S PR#ESENTENT	
882	2	VS	LES	COMPOS#ES	3 V		
885	3			PR#ESENTENT			
885	3	SP		PR#ESENTENT	4 V		
887	4		DES	LIAISONS			

PHRASE NO : 39 ELLE COMMENCE AU MOT NO: 6183

$00001420 ∂A CET EFFET , UN DOSSIER PORTANT SUR LES POINTS MENTIONN#ES AU PARAGRAPHE &I DE L' ARTICLE 1ER DU D#ECRET N. 61-1036 DU 13 SEPTEMBRE 1961 EST ADRESS#E AU DIRECTEUR D#EPARTEMENTAL DE LA CONSTRUCTION , QUI DISPOSE D' UN D#ELAI D' UN MOIS POUR PROC#EDER ∂A L' EXAMEN DU PROJET .

******* CHAINES VERBALES RECONNUES

PORTANT						S PORTANT	
6187	1	VS	UN	DOSSIER	2 V		
6188	2			PORTANT			COMPLEMENT ?
EST ADRESS#E						P ADRESS#E	
6187	1		UN	DOSSIER			
6209	6	SP		ADRESS#E	1 V		C. D'AGENT ?
6209	6			ADRESS#E			
6211	7	VC	AU	DIRECTEUR	6 V		
DISPOSE						S DISPOSE	
6211	6	VS	AU	DIRECTEUR	8 D		
6215	7	VS	DE LA	CONSTRUCTION	8 D		
6218	8			DISPOSE			COMPLEMENT ?

PHRASE NO : 56 ELLE COMMENCE AU MOT NO: 1577

TOUTEFOIS , CES DISPOSITIONS NE SERONT PAS OPPOSABLES AUX PERSONNES QUI AURONT D#EPOS#E UNE DEMANDE DE PERMIS DE CONSTRUIRE AVANT LE 1ER JUIN 1959 .

******* CHAINES VERBALES RECONNUES

NE SERONT PAS OPPOSABLES						P OPPOSABLES	
1580	1		CES	DISPOSITIONS			
1584	2	SP		OPPOSABLES	1 V		C. D'AGENT ?
1584	2			OPPOSABLES			
1586	3	VC	AUX	PERSONNES	2 V		
AURONT D#EPOS#E						S D#EPOS#E	
1586	2	VS	AUX	PERSONNES	3 V		
1589	3			D#EPOS#E			
1589	3	SP		D#EPOS#E	4 V		
1591	4		UNE	DEMANDE			

fig. 66. (a). Reconnaissance de liens anaphoriques. Les substitutions donnent lieu à des relations VS.

PHRASE NO : 34 ELLE COMMENCE AU MOT NO: 5900

LES SOCI#ET#ES DE CR#EDIT IMMOBILIER QUI SUSCITENT ET G@ERENT LES SOCI#ET#ES CIVILES CONSTITU#EES EN APPLICATION DU D#ECRET N. 59-700 DU 6 JUIN 1959 ET LES SOCI#ET#ES ANONYMES QUI ASSURENT LA R#EALISATION DES PROGRAMMES DE CONSTRUCTION DES SOCI#ET#ES CIVILES PEUVENT PERCEVOIR , AU TOTAL , UNE CONTRIBUTION INITIALE AU PLUS #EGALE @A 4 % DU PRIX DE REVIENT DES LOGEMENTS , TOUTES D#EPENSES CONFONDUES .

******* CHAINES VERBALES RECONNUES

	SUSCITENT					S SUSCITENT	
5901	1	VS		LES	SOCI#ET#ES	3 V	
5906	3				SUSCITENT		COMPLEMENT
	G@ERENT					S G@ERENT	
5901	1	VS		LES	SOCI#ET#ES	3 V	
5908	3				G@ERENT		
5908	3	SP			G@ERENT	4 V	
5910	4			LES	SOCI#ET#ES		
	ASSURENT					S ASSURENT	
5923	5	VS	DU	LES	SOCI#ET#ES	6 V	
5926	6				ASSURENT		
5926	6	SP			ASSURENT	7 V	
5928	7			LA	R#EALISATION		
	PEUVENT PERCEVOIR					S PERCEVOIR	
5923	5	VS	DU	LES	SOCI#ET#ES	10 V	
5937	10				PERCEVOIR		COMPLEMENT

fig. 66. (b). Reconnaissance de liens anaphoriques. Les substitutions donnent lieu à des relations VS.

PHRASE NO : 4 ELLE COMMENCE AU MOT NO: 82

CE RADIONUCL#EIDE EST SUSCEPTIBLE DE SE FIXER SUR LES PARTICULES EN SUSPENSION QUI SONT PI#EG#EES PAR LES ESP@ECES FILTREUSES TELLES QUE LES MOULES (&MYTILUS &EDULIS &L.) .

	SONT PI#EG#EES					P PI#EG#EES	
91	2		SUR	LES	PARTICULES		
96	4	SP			PI#EG#EES	2 V	
96	4				PI#EG#EES		
99	5	VS		LES	ESP@ECES	4 V	

PHRASE NO : 68 ELLE COMMENCE AU MOT NO: 1780

DEUX SECTIONS PERMANENTES PEUVENT ~ETRE APPEL#EES PAR LE MINISTRE OU PAR LE VICE-PR#ESIDENT @A SI#EGER R#EUNIES EN RAISON DE LA NATURE DES PROBL@EMES QUI LEUR SONT SOUMIS .

	SONT SOUMIS					P SOUMIS	
1800	5		DES		PROBL@EMES		
1804	6	SP			SOUMIS	5 V	C. D'AGENT

fig. 66. (c). Reconnaissance de liens anaphoriques. Les substitutions donnent lieu à des relations VD.

```
PHRASE NO :  1   ELLE COMMENCE AU MOT NO:  9403

$&F&B&J3&A1000&A&A&A TROIS S#ERIES DE DISPOSITIONS D#ETERMINENT LES OP#ERATIONS IMPOSABLES : IL S'
AGIT , TOUT D' ABORD , D'UNE MANI∂ERE G#EN#ERALE , DES AFFAIRES QUI REL∂EVENT D' UNE ACTIVIT#E DE
NATURE INDUSTRIELLE OU COMMERCIALE ET QUI , ∂A CE SEUL TITRE , ENTRENT DANS LE CHAMP D' APPLICATION
DE LA &T.&V.&A. .

REL∂EVENT                                                          S REL∂EVENT

9422  4  VS   DES             AFFAIRES                5 V
9424  5                       REL∂EVENT                                                     COMPLEMENT
------------------------------------------------------------------------------------------------------
ENTRENT                                                            S ENTRENT                SUJET ?
                                                                                            COMPLEMENT


PHRASE NO :  4   ELLE COMMENCE AU MOT NO:  9510

$&F&B&J3&A1110&A&A&A LES AFFAIRES * FAITES EN &FRANCE * VIS#EES ∂A L' ARTICLE 256 SONT LES
OP#ERATIONS QUE LES PERSONNES PHYSIQUES OU MORALES PEUVENT FAIRE DANS UN CADRE CONTRACTUEL ET QUI
PR#ESENTENT LES CARACT#ERISTIQUES DES ACTES DE COMMERCE .

PR#ESENTENT                                                        S PR#ESENTENT            SUJET ?

9540  7  SP                   PR#ESENTENT             8 V
9542  8             LES       CARACT#ERISTIQUES
------------------------------------------------------------------------------------------------------


PHRASE NO :114   ELLE COMMENCE AU MOT NO:  3761

$00000990 DANS LES PI∂ECES PRINCIPALES , LORSQUE TOUTES LES BAIES D' UN LOGEMENT OUVRENT SUR UNE
SEULE FA$CADE , DANS LES PI∂ECES DONT LES BAIES SONT #EQUIP#EES DE MANI∂ERE ∂A FERMER
HERM#ETIQUEMENT , LE RENOUVELLEMENT DE L' AIR DOIT -ETRE ASSUR#E DE FA$CON PERMANENTE PAR DES
OUVERTURES D' #EVACUATION DE L' AIR ET PAR DES OUVERTURES D' ENTR#EE D' AIR CONVENABLEMENT
DISPOS#EES ET AM#ENAG#EES , DE SECTION SUFFISANTE , QUI POURRONT -ETRE PR#ECIS#EES PAR DES
INSTRUCTIONS INTERMINIST#ERIELLES .

POURRONT -ETRE PR#ECIS#EES                                         P PR#ECIS#EES            COMPLEMENT

3829 16                       PR#ECIS#EES
3832 17  VS         DES       INSTRUCTIONS           16 V
------------------------------------------------------------------------------------------------------
```

fig. 67. (a). Silences totaux enregistrés lors de la reconnaissance des antécédents de l'anaphorique "qui".

```
PHRASE NO : 52   ELLE COMMENCE AU MOT NO:  1497

L' #ECHANGE NE PEUT -ETRE IMPOS#E AUX LOCATAIRES QUI ONT #ET#E SINISTR#E DANS LES LIEUX LOU#ES .

*******  CHAINES VERBALES RECONNUES

NE PEUT -ETRE IMPOS#E                                              P IMPOS#E

1498  1             L'        #ECHANGE
1502  2  SP                   IMPOS#E                 1 V                                   C. D'AGENT

1502  2                       IMPOS#E
1504  3  VC   AUX             LOCATAIRES              2 V
------------------------------------------------------------------------------------------------------
ONT #ET#E SINISTR#E                                                P SINISTR#E              COMPLEMENT
                                                                                            C. D'AGENT

1508  3                       SINISTR#E
1511  4  VC   DANS  LES       LIEUX                   3 V
------------------------------------------------------------------------------------------------------
```

fig. 67. (b). Silence total dû à une orthographication incorrecte. (SINISTRÉ au lieu de SINISTRÉS).

IV. 2. 5. Enchaînement des procédures :

Nous reprenons dans ce paragraphe, en le précisant, et sous une forme graphique, l'enchaînement des procédures que nous avons décrites dans le présent chapitre.

En effet, s'il y a enchaînement entre les procédures de segmentation en chaînes nominales et/ou verbales de la phrase d'une part, et les procédures de reconnaissance des relations de dépendance potentielles d'autre part, il n'y a pas véritablement un ordre strict à respecter quant au passage de ces dernières procédures, lors de la reconnaissance des relations homo ou hétérosyntagmatiques.

Autrement dit, les filtres homosyntagmatiques sont indépendants les uns des autres et les filtres hétérosyntagmatiques sont indépendants les uns des autres.

Ainsi, pour une chaîne verbale donnée, les reconnaissances des relations VS, VD ou VI_p sont, du moins dans cette étape, indépendantes les unes des autres.

Si donc il n'y a pas d'enchaînement particulier pour les relations homo ou hétérosyntagmatiques, il y a par contre enchaînement entre relations homosyntagmatiques d'une part, et hétérosyntagmatiques d'autre part.

En effet, nous avons vu, par exemple que l'élément sujet de la relation VS hétérosyntagmatique devait être libre.

Notons également qu'il y a un autre enchaînement, dynamique cette fois, établi par la phrase elle-même, qui est analysée de la gauche vers la droite. En particulier les chaînes verbales sont examinées séquentiellement de la gauche vers la droite.

Ceci permet alors, en rendant certains termes liés à une chaîne verbale donnée, d'éliminer d'emblée certains candidats pour des relations avec une chaîne verbale postérieure. C'est par exemple le cas du pronom relatif "qui" dans certaines configurations (voir Remarque p. 135).

L'enchaînement des traitements décrits dans le présent chapitre peut donc être représenté comme sur la figure 68.

V. Conclusion :

En guise de conclusion nous rappellerons que nous avons voulu dans ce chapitre exposer la composante syntaxique du système que nous proposons.

Cette composante syntaxique nous a permis ainsi d'obtenir, pour chaque phrase, une ou plusieurs représentations potentielles sous forme de dépendances.

Les pages suivantes offrent à cet égard quelques planches d'analyse de phrases extraites pour la plupart de textes sur la règlementation de l'urbanisme (cf. fig. 69).

Bien que nous ayions attribué l'adjectif "potentielles" à toutes les relations de dépendance que nous avons établies lors de cette étape, et au risque d'anticiper quelque peu sur le chapitre V, nous remarquerons que nous avons été, somme toute, rarement en situation d'ambiguïté (situation 2 du tableau fig. 23). C'est dire que l'hypothèse que nous émettrons au chapitre V, § I. 1. se trouve ainsi, d'ores et déjà, en partie vérifiée.

L'apprentissage de relations LS dans ces conditions est dès lors aisé et la résolution des ambiguïtés sémantiques possible comme nous le verrons dans le chapitre V.

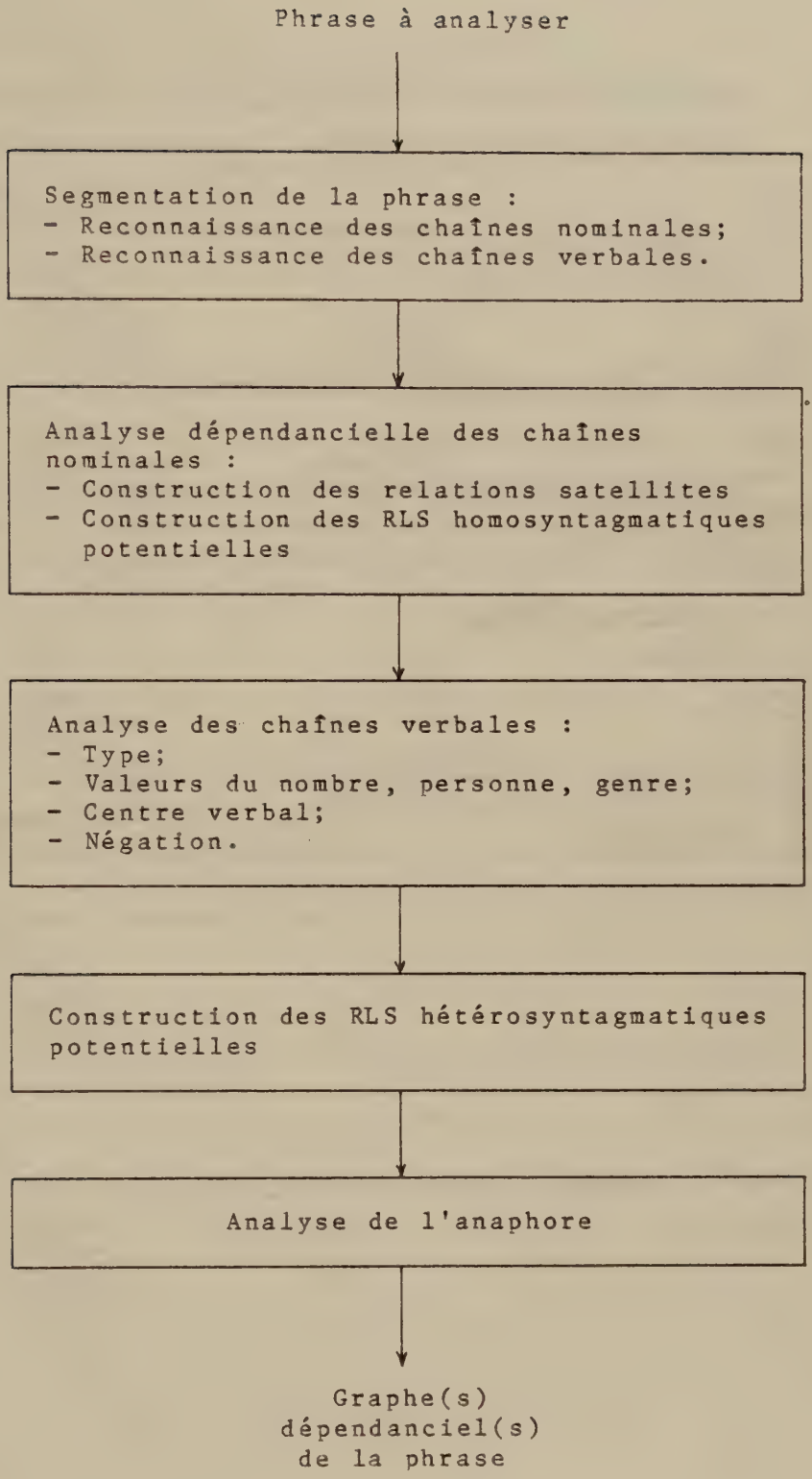


fig. 68 Enchaînement des procédures pour l'obtention des graphes de dépendance potentiels d'une phrase.

```
        PHRASE NO : 54   ELLE COMMENCE AU MOT NO:  1537

        UN NOUVEL ACTE DE LOCATION EST PASS#E PAR L' ORGANISME PROPRI#ETAIRE AVEC CHACUN DES INT#ERESS#ES
        POUR LES LOCAUX DANS LESQUELS ILS SONT INSTALL#ES .

.....

*******  CHAINES NOMINALES RECONNUES

        UN NOUVEL ACTE DE LOCATION

  1538  1  SA                      NOUVEL                  2 V
  1539  2                  UN      ACTE
  1541  3  SS   DE                 LOCATION                2 V
-------------------------------------------------------------------------------------------------------------------
        PAR L' ORGANISME PROPRI#ETAIRE

  1546  1                  L'      ORGANISME
  1547  2  SP                      PROPRI#ETAIRE           1 V
-------------------------------------------------------------------------------------------------------------------
        AVEC CHACUN DES INT#ERESS#ES
-------------------------------------------------------------------------------------------------------------------
        POUR LES LOCAUX (*)
-------------------------------------------------------------------------------------------------------------------
.....

*******  CHAINES VERBALES RECONNUES

        EST PASS#E                                                          P PASS#E

  1539  1                  UN      ACTE
  1543  3  SP                      PASS#E                  1 V

  1543  3                          PASS#E
  1546  4  VS                 L'      ORGANISME           3 V
  1551  5  VC   AVEC          DES     INT#ERESS#ES        3 V
-------------------------------------------------------------------------------------------------------------------
        SONT INSTALL#ES                                                     P INSTALL#ES
COMPL DE TYPE PRONOMINAL :  ILS

                                                                                        C. D'AGENT
-------------------------------------------------------------------------------------------------------------------
```

fig. 69. (a).

(*) Pour le moment, le système ne reconnaît qu'une relation VI_p par chaîne verbale.

```
        PHRASE NO : 59   ELLE COMMENCE AU MOT NO:  1633

        $00000110 LES MEMBRES DU CONSEIL SUP#RIEUR DU MINIST∅ERE DE LA CONSTRUCTION SONT NOMM#ES POUR
        DEUX ANS PAR ARR-ET#E DU MINISTRE DE LA CONSTRUCTION .

.....

*******  CHAINES NOMINALES RECONNUES

        LES MEMBRES DU CONSEIL SUP#RIEUR DU MINIST∅ERE DE LA CONSTRUCTION

  1635  1                LES       MEMBRES
  1637  2  SS   DU                 CONSEIL                1 V
  1638  3  SP                      SUP#RIEUR              2 V
  1640  4  SS   DU                 MINIST∅ERE             2 D    1 D
  1643  5  SS   DE       LA        CONSTRUCTION           4 D    2 D    1 D
-----------------------------------------------------------------------------

        POUR DEUX ANS PAR ARR-ET#E DU MINISTRE DE LA CONSTRUCTION

  1647  1        POUR              DEUX ANS
  1649  2                          ARR-ET#E
  1651  3  SS   DU                 MINISTRE               2 V
  1654  4  SS   DE       LA        CONSTRUCTION           3 D    2 D
-----------------------------------------------------------------------------

.....

*******  CHAINES VERBALES RECONNUES

        SONT NOMM#ES                                                  P NOMM#ES

  1635  1                LES       MEMBRES
  1645  5  SP                      NOMM#ES                1 V

  1645  5                          NOMM#ES
  1647  6  VC   POUR               DEUX ANS               5 V
  1649  7  VS                      ARR-ET#E               5 V
-----------------------------------------------------------------------------

        PHRASE NO : 70   ELLE COMMENCE AU MOT NO:  7404

        LE MINISTRE DE LA CONSTRUCTION PEUT D#ESIGNER DES RAPPORTEURS .

.....

*******  CHAINES NOMINALES RECONNUES

        LE MINISTRE DE LA CONSTRUCTION

  7405  1                LE        MINISTRE
  7408  2  SS   DE       LA        CONSTRUCTION           1 V
-----------------------------------------------------------------------------

        DES RAPPORTEURS
-----------------------------------------------------------------------------

.....

*******  CHAINES VERBALES RECONNUES

        PEUT D#ESIGNER                                                S D#ESIGNER

  7405  1  VS               LE        MINISTRE               3 V
  7410  3                             D#ESIGNER

  7410  3  SP                         D#ESIGNER              4 V
  7412  4                DES         RAPPORTEURS
-----------------------------------------------------------------------------
```

fig. 69. (b).

Interaction entre les traitements de ces deux phrases : voir chapitre V , § IV. 3., fig. 97 et 98 (a).

PHRASE NO : 68 ELLE COMMENCE AU MOT NO: 7343

$00001730 LA COMMISSION PEUT DEMANDER TOUTE ENQU~ETE OU SUPPL#EMENT D' INFORMATION QUI LUI PARA~ITRAIT N#ECESSAIRE ∂A L' INSTRUCTION DES AFFAIRES QUI LUI SONT SOUMISES .

****** CHAINES NOMINALES RECONNUES

LA COMMISSION

TOUTE ENQU~ETE

SUPPL#EMENT D' INFORMATION

7356	1				SUPPL#EMENT			
7358	2	SS	D'		INFORMATION	1 V		

∂A L' INSTRUCTION DES AFFAIRES

7365	1		∂A	L'	INSTRUCTION			
7367	2	SS	DES		AFFAIRES	1 V		

******* CHAINES VERBALES RECONNUES

PEUT DEMANDER — S DEMANDER

7350	1	VS		LA	COMMISSION	2 V			
7352	2				DEMANDER				
7352	2	SP			DEMANDER	3 V			
7354	3			TOUTE	ENQU~ETE				

PARA~ITRAIT N#ECESSAIRE — P N#ECESSAIRE

7356	3				SUPPL#EMENT				
7358	4		D'		INFORMATION				
7362	5	SP			N#ECESSAIRE	4 D	3 D		C. D'AGE
7362	5				N#ECESSAIRE				
7365	6	VC	∂A	L'	INSTRUCTION	5 V			

SONT SOUMISES — P SOUMISES

7367	6		DES		AFFAIRES				
7371	7	SP			SOUMISES	6 V			C. D'AGE

fig. 69. (c).

```
        PHRASE NO : 79   ELLE COMMENCE AU MOT NO:  7546

        SON PR#ESIDENT PR#EPARE LES CONVOCATIONS ET LES ORDRES DU JOUR ET #ETABLIT LES PROCƏES-VERBAUX .

.....

*******  CHAINES NOMINALES RECONNUES

        SON PR#ESIDENT
-----------------------------------------------------------------------------------------------------------------------
        LES CONVOCATIONS
-----------------------------------------------------------------------------------------------------------------------
        LES ORDRES DU JOUR

  7553  1                 LES       ORDRES
  7555  2  SS   OU                   JOUR                  1 V
-----------------------------------------------------------------------------------------------------------------------
        LES PROCƏES-VERBAUX
-----------------------------------------------------------------------------------------------------------------------
.....

*******  CHAINES VERBALES RECONNUES

        PR#EPARE                                                                    S PR#EPARE

  7547  1  VS              SON       PR#ESIDENT            2 V
  7548  2                            PR#EPARE

  7548  2  SP                         PR#EPARE              3 V
  7550  3                 LES       CONVOCATIONS
-----------------------------------------------------------------------------------------------------------------------
        #ETABLIT                                                                    S #ETABLIT

  7547  1  VS              SON       PR#ESIDENT            5 V
  7557  5                            #ETABLIT

  7557  5  SP                         #ETABLIT              6 V
  7559  6                 LES       PROCƏES-VERBAUX
=======================================================================================================================

        PHRASE NO : 80   ELLE COMMENCE AU MOT NO:  7561

        $00001780 LES AVIS #EMIS PAR LA COMMISSION SONT INSCRITS SUR UN REGISTRE ET SIGN#ES PAR LE
        PR#ESIDENT ET LE SECR#ETAIRE .

.....

*******  CHAINES NOMINALES RECONNUES

        LES AVIS #EMIS PAR LA COMMISSION

  7563  1                 LES       AVIS
  7564  2  SP                       #EMIS                 1 V
  7567  3  VS              LA        COMMISSION            2 V
-----------------------------------------------------------------------------------------------------------------------
        SUR UN REGISTRE
-----------------------------------------------------------------------------------------------------------------------
        PAR LE PR#ESIDENT
-----------------------------------------------------------------------------------------------------------------------
        LE SECR#ETAIRE
-----------------------------------------------------------------------------------------------------------------------
.....

*******  CHAINES VERBALES RECONNUES

        SONT INSCRITS                                                               P INSCRITS

  7563  1                 LES       AVIS
  7569  3  SP                       INSCRITS              1 V                                    C. D'AGENT ?

  7569  3                            INSCRITS
  7572  4  VC   SUR       UN        REGISTRE              3 V
-----------------------------------------------------------------------------------------------------------------------
        ET SIGN#ES                                                                  P SIGN#ES

  7563  1                 LES       AVIS
  7574  4  SP                       SIGN#ES               1 V

  7574  4                            SIGN#ES
  7577  5  VS              LE        PR#ESIDENT            4 V
-----------------------------------------------------------------------------------------------------------------------
```

fig. 69. (d).

PHRASE NO : 91 ELLE COMMENCE AU MOT NO: 13329

$&F&B&J3&A1110&A&M&A LES OP#ERATIONS DE COMMISSION ET DE COURTAGE CONSTITUENT UNE VARI#ET#E DE PRESTATIONS DE SERVICES QUI REV¬ETENT TOUTES LES DEUX UN CARACT@ERE COMMERCIAL .

******* CHAINES NOMINALES RECONNUES

LES OP#ERATIONS DE COMMISSION

13331	1		LES	OP#ERATIONS	
13333	2	SS DE		COMMISSION	1 V

DE COURTAGE

UNE VARI#ET#E DE PRESTATIONS DE SERVICES

13339	1		UNE	VARI#ET#E	
13341	2	SS DE		PRESTATIONS	1 V
13343	3	SS DE		SERVICES	2 V

UN CARACT@ERE COMMERCIAL

13350	1		UN	CARACT@ERE	
13351	2	SP		COMMERCIAL	1 V

******* CHAINES VERBALES RECONNUES

CONSTITUENT — S CONSTITUENT

13331	1	VS	LES	OP#ERATIONS	4 V
13337	4			CONSTITUENT	
13337	4	SP		CONSTITUENT	5 V
13339	5		UNE	VARI#ET#E	

REV¬ETENT — S REV¬ETENT

13341	5	VS DE		PRESTATIONS	7 D
13343	6	VS DE		SERVICES	7 D
13345	7			REV¬ETENT	
13345	7	SP		REV¬ETENT	8 V
13350	8		UN	CARACT@ERE	

fig. 69. (e)

PHRASE NO :100 ELLE COMMENCE AU MOT NO: 3346

$00000740 LE CARACT@ERE EXP#ERIMENTAL PEUT ¬ETRE RECONNU AUX CONSTRUCTIONS QUI ONT POUR OBJET D' ABAISSER LE PRIX DE LA CONSTRUCTION OU D' AM#ELIORER LES TECHNIQUES DU B¬ATIMENT .

******* CHAINES NOMINALES RECONNUES

LE CARACT@ERE EXP#ERIMENTAL

3348	1		LE	CARACT@ERE		
3349	2	SP		EXP#ERIMENTAL	1 V	

AUX CONSTRUCTIONS

LE PRIX DE LA CONSTRUCTION

3360	1		LE	PRIX		
3363	2	SS DE	LA	CONSTRUCTION	1 V	

LES TECHNIQUES DU B¬ATIMENT

3368	1		LES	TECHNIQUES		
3370	2	SS DU		B¬ATIMENT	1 V	

******* CHAINES VERBALES RECONNUES

PEUT ¬ETRE RECONNU — P RECONNU

3348	1		LE	CARACT@ERE		
3352	2	SP		RECONNU	1 V	C. D'AGENT ?
3352	2			RECONNU		
3354	3	VC AUX		CONSTRUCTIONS	2 V	

ONT POUR OBJET D' ABAISSER — S ABAISSER

3354	2	VS AUX		CONSTRUCTIONS	3 V	
3358	3			ABAISSER		
3358	3	SP		ABAISSER	4 V	
3360	4		LE	PRIX		

OU D' AM#ELIORER — S AM#ELIORER — SUJET ?

3366	5	SP		AM#ELIORER	6 V	
3368	6		LES	TECHNIQUES		

fig. 69. (f).

```
        PHRASE NO :126   ELLE COMMENCE AU MOT NO:  4228

        $00001060 LORSQUE LES CONDUITS SONT CON$CUS DE TELLE MANIØERE QU' ILS NE PEUVENT ASSURER SEULS LEUR
        STABILIT#E , ILS DOIVENT ¬ETRE ADOSS#ES ØA DES #EL#EMENTS PORTEURS CONSTRUITS EN MAT#ERIAUX
        INCOMBUSTIBLES D' UNE STABILIT#E ET D' UNE TENUE AU FEU SUFFISANTES .

.....

*******  CHAINES NOMINALES RECONNUES

        LES CONDUITS
-------------------------------------------------------------------------------------------------------------------------------
        LEUR STABILIT#E
-------------------------------------------------------------------------------------------------------------------------------
        ØA DES #EL#EMENTS PORTEURS CONSTRUITS EN MAT#ERIAUX INCOMBUSTIBLES D' UNE STABILIT#E

  4250  1        ØA        DES       #EL#EMENTS
  4251  2  SP                         PORTEURS                  1 V
  4252  3  SP                         CONSTRUITS                1 V
  4254  4  VC   EN                     MAT#ERIAUX                3 V
  4255  5  SP                         INCOMBUSTIBLES            4 V
  4258  6  SS   D'        UNE         STABILIT#E                4 V
-------------------------------------------------------------------------------------------------------------------------------
        ET D' UNE TENUE AU FEU SUFFISANTES

  4262  1        D'        UNE       TENUE
  4264  2  SS   AU                     FEU                       1 V
  4265  3                              SUFFISANTES
-------------------------------------------------------------------------------------------------------------------------------
.....

*******  CHAINES VERBALES RECONNUES

        SONT CON$CUS                                                          P CON$CUS

  4231  1                    LES       CONDUITS
  4233  2  SP                         CON$CUS                   1 V                                    C. D'AGENT ?
-------------------------------------------------------------------------------------------------------------------------------
        NE PEUVENT ASSURER SEULS                                              S ASSURER
SUJET DE TYPE PRONOMINAL :  ILS


  4239  2  SP                         ASSURER                   3 V
  4242  3                    LEUR      STABILIT#E
-------------------------------------------------------------------------------------------------------------------------------
        DOIVENT ¬ETRE ADOSS#ES                                                P ADOSS#ES
COMPL DE TYPE PRONOMINAL :  ILS

                                                                                                             C. D'AGENT ?
  4247  3                              ADOSS#ES
  4250  4  VC   ØA        DES         #EL#EMENTS                3 V
-------------------------------------------------------------------------------------------------------------------------------
```

fig. 69. (g).

PHRASE NO :130 ELLE COMMENCE AU MOT NO: 4304

LES FACES DIRECTEMENT ACCESSIBLES DES CONDUITS ADOSS#ES ∂A UN MUR EXT#ERIEUR OU ∂A UNE PAROI DE CAGE D' ESCALIER , OU INCORPOR#ES DANS CES PAROIS , DOIVENT AVOIR UNE ISOLATION SUFFISANTE POUR QUE LE REFROIDISSEMENT NE CONTRARIE PAS LE TIRAGE .

******* CHAINES NOMINALES RECONNUES

LES FACES DIRECTEMENT ACCESSIBLES DES CONDUITS ADOSS#ES ∂A UN MUR EXT#ERIEUR

4305	1			LES	FACES	
4307	2	SP			ACCESSIBLES	1 V
4309	3	SS	DES		CONDUITS	1 V
4310	4	SP			ADOSS#ES	3 V
4313	5	VC	∂A	UN	MUR	4 V
4314	6	SP			EXT#ERIEUR	5 V

OU ∂A UNE PAROI DE CAGE D' ESCALIER

4318	1		∂A	UNE	PAROI	
4320	2	SS	DE		CAGE	1 V
4322	3	SS	D'		ESCALIER	2 V

DANS CES PAROIS

UNE ISOLATION SUFFISANTE

4333	1			UNE	ISOLATION	
4334	2	SP			SUFFISANTE	1 V

LE REFROIDISSEMENT

LE TIRAGE

******* CHAINES VERBALES RECONNUES

		DOIVENT AVOIR					S AVOIR
4305	1	VS		LES	FACES	8 V	
4331	8				AVOIR		
4331	8	SP			AVOIR	9 V	
4333	9			UNE	ISOLATION		
		NE CONTRARIE PAS					S CONTRARIE
4337	9	VS		LE	REFROIDISSEMENT	10 V	
4339	10				CONTRARIE		
4339	10	SP			CONTRARIE	11 V	
4342	11			LE	TIRAGE		

fig. 69. (h)

```
PHRASE NO :134   ELLE COMMENCE AU MOT NO:  4442

$00001240 LES ATTRIBUTIONS DE LOGEMENTS DANS LES IMMEUBLES VIS#ES ∂A L' ARTICLE 1ER SONT FAITES PAR
L' ORGANISME D' HABITATIONS ∂A LOYER MOD#ER#E EN ACCORD AVEC LA COLLECTIVIT#E OU L' ORGANISME DE
S#ECURIT#E SOCIALE QUI A GARANTI LE D#EFICIT D' EXPLOITATION DE CES LOGEMENTS LORSQU' UNE TELLE
GARANTIE A #ET#E DONN#EE .

.....

******* CHAINES NOMINALES RECONNUES

        LES ATTRIBUTIONS DE LOGEMENTS DANS LES IMMEUBLES VIS#ES ∂A L' ARTICLE 1ER

 4444  1                    LES       ATTRIBUTIONS
 4446  2  SS   DE                     LOGEMENTS          1 V
 4449  3       DANS         LES       IMMEUBLES
 4450  4  SP                          VIS#ES             3 V
 4453  5  VC   ∂A           L'        ARTICLE            4 V
 4454  6  SP                          1ER                5 V
-------------------------------------------------------------------------------------------------------
        PAR L' ORGANISME D' HABITATIONS ∂A LOYER MOD#ER#E

 4459  1                    L'        ORGANISME
 4461  2  SS   D'                     HABITATIONS        1 V
 4463  3  SS   ∂A                     LOYER              2 V
 4464  4  SP                          MOD#ER#E           3 U ***    1 D
-------------------------------------------------------------------------------------------------------
        EN ACCORD AVEC LA COLLECTIVIT#E
-------------------------------------------------------------------------------------------------------
        L' ORGANISME DE S#ECURIT#E SOCIALE

 4470  1                    L'        ORGANISME
 4472  2  SS   DE                     S#ECURIT#E         1 V
 4473  3  SP                          SOCIALE            2 V
-------------------------------------------------------------------------------------------------------
        LE D#EFICIT D' EXPLOITATION DE CES LOGEMENTS

 4478  1                    LE        D#EFICIT
 4480  2  SS   D'                     EXPLOITATION       1 V
 4483  3  SS   DE           CES       LOGEMENTS          2 D   1 D
-------------------------------------------------------------------------------------------------------
        UNE TELLE GARANTIE

 4486  1  SA                          TELLE              2 V
 4487  2                    UNE       GARANTIE
-------------------------------------------------------------------------------------------------------

.....

******* CHAINES VERBALES RECONNUES

        SONT FAITES                                                          P FAITES

 4444  1                    LES       ATTRIBUTIONS
 4456  5  SP                          FAITES             1 V

 4456  5                              FAITES
 4459  6  VS                 L'        ORGANISME         5 V
 4467  9  VC   EN ACCORD LA           COLLECTIVIT#E      5 V
-------------------------------------------------------------------------------------------------------
        A GARANTI                                                            S GARANTI

 4470  9  VS                 L'        ORGANISME        11 D
 4472 10  VS   DE                     S#ECURIT#E       11 D
 4476 11                              GARANTI

 4476 11  SP                          GARANTI           12 V
 4478 12                    LE        D#EFICIT
-------------------------------------------------------------------------------------------------------
        A #ET#E DONN#EE                                                      P DONN#EE

 4487 14                    UNE       GARANTIE
 4490 15  SP                          DONN#EE           14 V                                  C. D'AGENT ?
-------------------------------------------------------------------------------------------------------
```

fig. 69. (i).

Chapitre quatrième

COMPOSANTE MORPHOLOGIQUE DU SYSTEME

CONSTRUCTION AUTOMATIQUE DE FAMILLES DE MOTS

DESCRIPTION PARADIGMATIQUE DE LA PHRASE

Introduction :

Dans le chapitre précédent nous avons décrit la composante syntaxique du système qui devait nous conduire aux diverses descriptions syntagmatiques potentielles de la phrase.

Nous allons dans ce chapitre décrire la composante morphologique du système qui nous conduira cette fois aux diverses descriptions paradigmatiques des éléments de la phrase, par abus de langage, nous dirons de la phrase.

C'est, comme nous le verrons dans le chapitre suivant , sur cette double description, syntagmatique et paradigmatique, qu'est fondée la représentation de la phrase, et plus généralement du contenu d'un texte, c'est à dire enfin de la connaissance.

C'est également la combinaison de ces deux descriptions qui rendra possible la détection de la synonymie lexico-syntaxique qui est, comme nous l'avons souligné, fondamentale si l'on veut accroître les capacités de résolution du système lors de la reconnaissance des relations de dépendance (cf. ch. II, § VI, ch. V, § III).

Il s'agit donc ici d'associer aux différents éléments de la phrase, les différents éléments du vocabulaire qui peuvent s'y substituer. La description qui en découlera sera cependant, comme précédemment, potentielle puisqu'elle ne fera pas intervenir les différents rééquilibrages syntaxiques ou sémantiques qui accompagnent généralement ces substitutions lexicales. C'est en effet dans le chapitre suivant que nous verrons comment interagissent et s'équilibrent les transformations lexicales et syntaxiques. Nous plaçant en régime d'analyse et non de synthèse, nous verrons en particulier que ces rééquilibrages pourront se faire dynamiquement, c'est à dire automatiquement (cf. ch. V, § III et IV).

Cette description paradigmatique potentielle, nous dirons aussi intrinsèque, suppose donc la connaissance des relations sémantiques qui peuvent exister entre les éléments du vocabulaire, telles par exemple les relations de synonymie ou d'antonymie, et plus généralement la connaissance des éléments du vocabulaire ayant une parenté sémantique plus ou moins accusée (cf. ch. II, § I et II).

Or, outre l'utilisation de synonymies du type "voiture - automobile" ou de l'antonymie du type "lenteur - rapidité", "simplicité - complexité", où l'on peut remarquer l'absence de toute proximité morphologique, les transfomations lexico-syntaxiques semblent se faire surtout au moyen de variations morphologiques.

En effet, l'étude de la formation des mots nouveaux montre qu'une langue enrichit son vocabulaire non seulement par des emprunts faits aux langues étrangères, mais encore et surtout, par la création de mots nouveaux au moyen de la dérivation et de la composition. (Notons que l'on distingue en général deux autres procédés de création, mais moins productifs : la formation par onomatopées et l'abréviation) (GREVISSE, 62) .

C'est ainsi, par exemple, que l'on peut obtenir par le jeu de la préfixation et/ou de la suffixation les mots suivants à partir du radical TAX :

TAXATION , TAXE , TAXER
DETAXATION , DETAXE , DETAXER
SURTAXATION , SURTAXE , SURTAXER

Le présent chapitre est précisément consacré à l'acquisition automatique de telles informations ("ensembles" ou "classes de mots" en l'occurrence) qui, souvent donc, sont à la base des substitutions lexicales dans la formation des paraphrases(*).

La description paradigmatique de la phrase sera par suite fondée sur la détermination automatique de ces classes de mots qui, finalement, correspondent à ce que l'on appelle traditionnellement "familles de mots".

Par abus de langage, nous appellerons "concepts" ces familles de mots.

Nous allons exposer dans ce qui suit un algorithme permettant d'obtenir, à partir d'un simple lexique de mots (chaque mot est accompagné dans le lexique de sa ou ses catégories grammaticales hors contexte), des règles permettant la construction de familles de mots.

(*) *1° Interdiction d'afficher sur les murs.*

2° Affichage mural interdit.

3° L'affichage sur les murs est interdit.

etc..

Les éléments de ces différentes phrases présentent une certaine proximité morphologique. Ce sont ces proximités morphologiques que nous nous proposons d'appréhender automatiquement.

II. Notion de famille de mots•:

M. GREVISSE (62) définit ainsi la famille de mots :
"C'est l'ensemble des mots qui peuvent se grouper autour d'un radical commun d'où ils ont été tirés par la dérivation ou par la composition"

Avant de faire nôtre une telle définition, il convient de noter les divers critères qui peuvent présider à la définition de la notion de famille de mots, compte tenu des multiples applications où cette notion peut intervenir. La question qui se pose, en effet, est de savoir à quelles conditions deux mots donnés peuvent être rapportés à un même ensemble de mots.

D'une façon très générale, si le but est d'engendrer ou de reconnaître des phrases équivalentes du point de vue du sens, on pourrait dire que deux mots, de même origine morphologique ou non, appartiennent à un même ensemble sémantique si, quelle que soit une phrase S_1 utilisant l'un des deux mots, on peut trouver une phrase S_2 "équivalente" utilisant l'autre mot. Toutefois la notion de véritable paraphrase, comme le faisait remarquer indirectement LA BRUYERE(*), peut toujours être combattue.

Par exemple :

a) *Traitement des banques de données textuelles et plus particulièrement juridiques ;*
b) *Traitement des banques de données juridiques et plus généralement textuelles.*

sont, selon le niveau où l'on a décidé d'arrêter l'analyse, équivalentes, pour la recherche documentaire par exemple, mais différentes quant à leur présupposés (DUCROT, 55) et leur impact psychologique.

En conséquence, dans ce qui suit, les équivalences de sens des mots appartenant à une même famille, seront examinés d'un point de vue pratique orienté vers la détection de la synonymie lexico-syntaxique (au sens large du terme), même si d'autres applications peuvent être envisagées.

On peut remarquer cependant que les deux approches, l'une basée sur l'étymologie (22) , l'autre sur la construction de phrases, équivalentes du point

---

(*)"Entre toutes les différentes expressions qui peuvent rendre une seule de nos pensées, il n'y en a qu'une qui soit la bonne" LA BRUYERE" (cité par M. GREVISSE (62, p. 108)).

de vue du sens, déterminent des ensembles de mots qui présentent une forte analogie : la dérivation et la composition conservent en général le sens, même si elles s'accompagnent d'un changement du statut syntaxique.

Dans le but de construire automatiquement ces familles de mots, nous avons développé une méthodologie particulière utilisant les propriétés morphologiques des mots ainsi que leurs valeurs grammaticales.

III. Principe de la méthode proposée pour l'élaboration automatique des familles de mots :

La méthode utilisée est basée sur le découpage d'un mot en :

chaîne préfixale + corps + chaîne suffixale (*)

On est alors confronté au problème classique de la validité des découpages obtenus.

Pour résoudre ce problème, nous développons une heuristique fondée sur la compatibilité des suffixes ou des préfixes ainsi dégagés.

IV. Reconnaissance automatique de familles de mots :

La méthode proposée consiste à engendrer des familles de mots à partir d'un dictionnaire. Celui-ci renferme, ainsi que nous pouvons le voir sur la figure 70 les mots accompagnés de leurs valeurs grammaticales hors contexte (**).

Notons que cette information associée à chaque mot permet d'éliminer d'ores et déjà un ensemble de mots dits outils (tels que les prépositions, conjonctions articles, pronoms, ...) qui ne peuvent participer à des familles de mots. Leur présence conduirait d'ailleurs à des familles fausses telles que < POUR, POURCEAU > , et qui, pour certaines d'entre elles, ne pourraient être éliminées autrement que par des moyens manuels du fait de l'existence de familles de mots similaires mais correctes comme < LION, LIONCEAU > .

(*) Nous utiliserons le symbole + pour noter la concaténation.

(**) Pour ce chapitre, nous avons retenu quatre parties du discours : substantif (SUBS), verbe (VRB), adjectif (ADJ) et adverbe (ADV) dénommées valeurs grammaticales et une cinquième valeur "AUTRE" regroupant tous les mots n'appartenant pas à l'une au moins des quatre classes précitées.

IV. 1. Familles de mots primitives :

Soient :

M l'ensemble des mots de la langue française **F**

P un ensemble de préfixes associés à **F**

S un ensemble de suffixes associés à **F**

et soit r une chaîne de caractères

Un sous-ensemble $F(r) \subset M$ est appelé famille de mots primitive définie par la chaîne r si ce sous-ensemble est déterminé comme suit :

$$F(r) = \langle m_i \mid m_i = p_i + r + s_i \text{ et } V(m_i) \cap V(s_i) \neq \emptyset,\ m_i \in \mathbf{M},\ p_i \in \mathbf{P},\ s_i \in \mathbf{S} \rangle$$

où $V(m_i)$ est l'ensemble des valeurs grammaticales hors contexte du mot m_i et $V(s_i)$ celui du suffixe s_i (par exemple V(AIN) = < ADJECTIF, SUBSTANTIF >)

Par abus de langage, nous appellerons "radical" la chaîne de caractères r.

Exemples :

F (HABIT) = < HABIT, HABIT + ABLE, HABIT + TANT, HABIT + AT, HABIT + ATION >

F (CENS) = < EN + CENS + ER, RE + CENS + ER, CENS + URE, ... >

Notons qu'un mot m donné peut appartenir à plusieurs familles primitives.

Exemples :

F (ART) = < ART, ART + ERE, ART + ISTE >
F (ARTIST) = < ARTIST + E, ARTIST + IQUE >

Ici ARTISTE ∈ F (ART) et F (ARTIST) mais F (ART) ≠ F (ARTIST)

F (ASSASS) = < ASSASS + IN, ASSASS + INER >
F (ASSASSIN) = < ASSASSIN, ASSASSIN + AT, ASSASSIN + ER >

dans le deuxième exemple F (ASSASS) ⊂ F (ASSASSIN)

Nous allons, pour la suite de l'exposé, considérer seulement les familles de mots primitives engendrées par les suffixes. Nous verrons plus loin (cf. § V) la manière dont on fera intervenir les préfixes. On considèrera donc dans la suite la famille primitive définie comme suit :

$$F(r) = \langle m_i \mid m_i = r + s_i \text{ et } V(m_i) \cap V(s_i) \neq \emptyset \quad m_i \in M, s_i \in S \rangle$$

Une telle famille sera dite suffixale.

La condition $V(m_i) \cap V(s_i) \neq \emptyset$ introduit lors de la décomposition des mots, une contrainte basée sur la compatibilité des valeurs grammaticales des mots et des suffixes (fig. 73). Cela permet alors d'éviter la création de familles primitives illicites que nous aurions eu à éliminer par la suite.

Exemples :

Considérons les suffixes E et RE avec :

$$V(E) = \langle SUBSTANTIF \rangle \text{ et } V(RE) = \langle VERBE \rangle$$

et soit le mot m = MONTRE avec

$$V(MONTRE) = \langle SUBSTANTIF \rangle$$

alors la seule décomposition possible de m est :

$$MONTRE = MONTR + E$$

L'interdiction de MONTRE = MONT + RE élimine des familles primitives, des occurrences telles que :

(MONTRE, MONTRER) , (MONTRE, MONTICULE), etc...

Remarque :

a) Pour tenir compte de ce que les linguistes appellent dérivation régressive ou rétrograde, nous avons adjoint à l'ensemble **S** des suffixes utilisés le suffixe "vide" que nous noterons b.

Exemples :

ACCORDER - ACCORD
GALOPER - GALOP

b) Nous avons également introduit les suffixes BLE et BILITE pour tenir compte des suffixes ABILITE, IBILITE et UBILITE. Ce choix est en fait peu adéquat et il eût été préférable d'introduire directement ces trois suffixes (cf. fig. 72).

	Mot		V(mot)			Mot		V(mot)	
1	¬ACRE	===	ADJ		61	ABORD	===	SUBS	
2	¬ACRET#E	===	SUBS		62	ABORDER	===	VRB	
3	¬AFFUT	===	SUBS		63	ABOUCHER	===	VRB	
4	¬AGE	===	SUBS		64	ABOUTIR	===	VRB	
5	¬AG#E	===	ADJ		65	ABOUTISSEMENT	===	SUBS	
6	¬AME	===	SUBS		66	ABOYER	===	VRB	
7	¬ANE	===	SUBS		67	ABRACADABRANT	===	ADJ	
8	¬ANERIE	===	SUBS		68	ABRASIF	===	ADJ	
9	¬ANESSE	===	SUBS		69	ABREUVER	===	VRB	
10	¬ANIER	===	SUBS		70	ABREUVOIR	===	SUBS	
11	¬ANON	===	SUBS		71	ABR#EGER	===	VRB	
12	¬APRE	===	ADJ		72	ABR#EG#E	===	ADJ	SUBS
13	¬APRET#E	===	SUBS		73	ABR#EVIATION	===	SUBS	
14	¬ATRE	===	SUBS		74	ABRI	===	SUBS	
15	¬ETRE	===	SUBS	VRB	75	ABRICOT	===	SUBS	
16	¬ILE	===	SUBS		76	ABRITER	===	VRB	
17	¬ILOT	===	SUBS		77	ABROGATION	===	SUBS	
18	%	===	SUBS		78	ABROGER	===	VRB	
19	A<IEUL	===	SUBS		79	ABRUPT	===	ADJ	
20	A¬INESSE	===	SUBS		80	ABRUTIR	===	VRB	
21	A¬IN#E	===	SUBS		81	ABSCONS	===	ADJ	
22	AB¬IME	===	SUBS		82	ABSENCE	===	SUBS	
23	AB¬IMER	===	VRB		83	ABSENT	===	ADJ	SUBS
24	ABAISSEMENT	===	SUBS		84	ABSENTER	===	VRB	
25	ABAISSER	===	VRB		85	ABSIDE	===	SUBS	
26	ABANDON	===	SUBS		86	ABSINTHE	===	SUBS	
27	ABANDONNER	===	VRB		87	ABSOLU	===	ADJ	
28	ABAQUE	===	SUBS		88	ABSOLUMENT	===	ADV	
29	ABASOURDIR	===	VRB		89	ABSOLUTION	===	SUBS	
30	ABAT	===	SUBS		90	ABSORBANT	===	ADJ	
31	ABATTAGE	===	SUBS		91	ABSORBER	===	VRB	
32	ABATTEMENT	===	SUBS		92	ABSORPTION	===	SUBS	
33	ABATTOIR	===	SUBS		93	ABSOUDRE	===	VRB	
34	ABATTRE	===	VRB		94	ABSOUT	===	ADJ	
35	ABBAYE	===	SUBS		95	ABSTENIR	===	VRB	
36	ABBESSE	===	SUBS		96	ABSTENTION	===	SUBS	
37	ABB#E	===	SUBS		97	ABSTINENCE	===	SUBS	
38	ABC@ES	===	SUBS		98	ABSTRACTION	===	SUBS	
39	ABC#EDER	===	VRB		99	ABSTRAIT	===	ADJ	
40	ABDIQUER	===	VRB		100	ABSURDE	===	ADJ	
41	ABDOMEN	===	SUBS		101	ABSURDIT#E	===	SUBS	
42	ABEILLE	===	SUBS		102	ABUS	===	SUBS	
43	ABERRATION	===	SUBS		103	ABUSER	===	VRB	
44	ABHORRER	===	VRB		104	ABUSIF	===	ADJ	
45	ABJECT	===	ADJ		105	ABUSIVEMENT	===	ADV	
46	ABJURER	===	VRB		106	ACABIT	===	SUBS	
47	ABLATION	===	SUBS		107	ACACIA	===	SUBS	
48	ABLETTE	===	SUBS		108	ACAD#EMICIEN	===	SUBS	
49	ABLUTION	===	SUBS		109	ACAD#EMIE	===	SUBS	
50	ABN#EGATION	===	SUBS		110	ACAJOU	===	SUBS	
51	ABOI	===	SUBS		111	ACANTHE	===	SUBS	
52	ABOIEMENT	===	SUBS		112	ACARI¬ATRE	===	ADJ	
53	ABOLIR	===	VRB		113	ACCABLER	===	VRB	
54	ABOLITION	===	SUBS		114	ACCALMIE	===	SUBS	
55	ABOMINABLE	===	ADJ		115	ACCAPARER	===	VRB	
56	ABONDANCE	===	SUBS		116	ACCENT	===	SUBS	
57	ABONDANT	===	ADJ		117	ACCENTUATION	===	SUBS	
58	ABONDER	===	VRB		118	ACCENTUER	===	VRB	
59	ABONNEMENT	===	SUBS		119	ACCEPTATION	===	SUBS	
60	ABONNER	===	VRB		120	ACCEPTER	===	VRB	

fig. 70. a. Extraits du dictionnaire ayant servi à l'expérience.

AMYGDALE	INEXTRICABLE	ASSIMILABLE	IMP#EN#ETRABLE
SCANDALE	PRATICABLE	MOUILLABLE	CURABLE
SANDALE	IMPRATICABLE	IN#EBRANLABLE	INCURABLE
VANDALE	VOCABLE	INVIOLABLE	DURABLE
C#ER#EALE	R#EVOCABLE	INCONSOLABLE	SECOURABLE
RAFALE	IRR#EVOCABLE	INCALCULABLE	INCOMMENSURABLE
GALE	CONSOLIDABLE	CUMULABLE	OBTURABLE
ASTRAGALE	FORMIDABLE	ANNULABLE	LIVRABLE
R#EGALE	INSONDABLE	BL¬AMABLE	OUVRABLE
CIGALE	INABORDABLE	AIMABLE	IRRECOUVRABLE
FRINGALE	DOMMAGEABLE	SUBLIMABLE	SABLE
MARTINGALE	AM#ENAGEABLE	INEXPRIMABLE	R#EALISABLE
HALE	PARTAGEABLE	INFLAMMABLE	IRR#EALISABLE
ENC#EPHALE	N#EGLIGEABLE	CONSOMMABLE	NORMALISABLE
PROVINCIALE	DIRIGEABLE	TRANSFORMABLE	UTILISABLE
FILIALE	INTERCHANGEABLE	CONVENABLE	R#EUTILISABLE
MERCURIALE	MALL#EABLE	INALI#ENABLE	INUTILISABLE
D#ECIMALE	PERM#EABLE	INEXPUGNABLE	INDEMNISABLE
ANORMALE	IMPERM#EABLE	MINABLE	IND#EFRISABLE
BACCHANALE	AGR#EABLE	ABOMINABLE	REVALORISABLE
FINALE	D#ESAGR#EABLE	INTERMINABLE	IN#EPUISABLE
BIENNALE	FABLE	IMPRESSIONNABLE	R#EVISABLE
DIAGONALE	AFFABLE	SUBVENTIONNABLE	INDISPENSABLE
M#ERIDIONALE	INFATIGABLE	RAISONNABLE	RESPONSABLE
R#EGIONALE	NAVIGABLE	INCUMABLE	IMPOSABLE
PALE	D#ETACHABLE	CAPABLE	OPPOSABLE
S#EPALE	IRR#EPROCHABLE	INCAPABLE	INOPPOSABLE
OPALE	INAPPR#ECIABLE	PALPABLE	REMBOURSABLE
CATH#EDRALE	PR#EJUDICIABLE	IMPALPABLE	INCASSABLE
G#EN#ERALE	JUSTICIABLE	COUPABLE	INLASSABLE
SPIRALE	N#EGOCIABLE	ARABLE	PASSABLE
CHORALE	SOCIABLE	IRR#EPARABLE	M#ECONNAISSABLE
MORALE	DISSOCIABLE	S#EPARABLE	P#ERISSABLE
CENTRALE	INDISSOCIABLE	INS#EPARABLE	IMP#ERISSABLE
SALE	DIABLE	COMPARABLE	INSAISISSABLE
TRANSVERSALE	LIQU#EFIABLE	INCOMPARABLE	AMORTISSABLE
SUCCURSALE	INQUALIFIABLE	INNOMBRABLE	INEXCUSABLE
VASSALE	PANIFIABLE	EX#ECRABLE	TABLE
#ETALE	BONIFIABLE	#ERABLE	RESPECTABLE
P#ETALE	INCONCILIABLE	CONSID#ERABLE	IN#ELUCTABLE
DIGITALE	R#ESILIABLE	IMPOND#ERABLE	RETABLE
CAPITALE	AMIABLE	PR#EF#ERABLE	#ETABLE
ORIENTALE	MANIABLE	TRANSF#ERABLE	INTRAITABLE
SENTIMENTALE	IND#ENIABLE	V#EN#ERABLE	HABITABLE
CROTALE	VARIABLE	VULN#ERABLE	INHABITABLE
VESTALE	INVARIABLE	R#EMUN#ERABLE	INDUBITABLE
SQUALE	FRIABLE	R#ECUP#ERABLE	PROFITABLE
CAVALE	INSATIABLE	MIS#ERABLE	CHARITABLE
RIVALE	VIABLE	INALT#ERABLE	V#ERITABLE
OVALE	SERVIABLE	IND#ECHIFFRABLE	IRRITABLE
C¬ABLE	JABLE	ADMIRABLE	#EQUITABLE
R¬ABLE	PR#EALABLE	D#ESIRABLE	IN#EVITABLE
PROBABLE	VALABLE	IND#ESIRABLE	#EPOUVANTABLE
IMPROBABLE	SEMBLABLE	D#EPLORABLE	ORIENTABLE
IMPERTURBABLE	VRAISEMBLABLE	M#EMORABLE	LAMENTABLE
CABLE	DISSEMBLABLE	HONORABLE	RENTABLE
IMPLACABLE	CYCLABLE	FAVORABLE	NOTABLE
IMPECCABLE	RENOUVELABLE	D#EFAVORABLE	POTABLE
APPLICABLE	INEFFLABLE	INEXORABLE	ADAPTABLE
INAPPLICABLE	R#EGLABLE	IN#ENARRABLE	INACCEPTABLE

VIADUC	BROCARD	P#ETARD	MOZARABE
AQUEDUC	DARD	R#ETARD	CRABE
ARCHIDUC	STANDARD	VANTARD	#EPHOEBE
OL#ECDUC	#ETENDARD	MOUTARD	GLOEBE
GAZODUC	SOUDARD	COUARD	PLOEBE
BOUC	FARD	JACQUARD	BRIBE
CAOUTCHOUC	CAFARD	BAVARD	SCRIBE
TRUC	BLAFARD	BOULEVARD	DIATRIBE
SUC	HAGARD	BUVARD	GALBE
STUC	REGARD	YARD	BULBE
WC	RINGARD	BOYARD	INGAMBE
RAD	RICHARD	FUYARD	IAMBE
PIED	MOUCHARD	L#EZARD	JAMBE
MARCHEPIED	BO<IARD	BORD	DITHYRAMBE
TR#EPIED	LIARD	ABORD	LIMBE
BLED	MILLIARD	B¬ABORD	NIMBE

fig. 70. b. Extrait du dictionnaire classé selon l'ordre alphabétique des mots miroirs.

1	A
2	AC
3	AD
4	AF
5	AG
6	AL
7	AN
8	ANT#E
9	ANTI
10	AP
11	AR
12	AS
13	AT
14	BI
15	BIEN
16	BIS
17	CIRCOM
18	CIRCON
19	CIRCUM
20	CIS
21	CO
22	COL
23	COM
24	CON
25	CONTRE
26	COR
27	D#E
28	D#ES
29	DIS
30	EM
31	EN
32	ENTR
33	ENTRE
34	EX
35	EXTRA
36	FOR
37	IL
38	IM
39	IN
40	INTER
41	INTRA
42	IR
43	JUXTA
44	MAL
45	MAU
46	M#E
47	M#ES
48	NON
49	PAR
50	P#EN
51	P#EN#E
52	POST
53	POUR
54	PR#E
55	PRO
56	QUADRI
57	R
58	RA
59	RE
60	RES
61	R#E
62	SOU
63	SUB
64	SUPER
65	SUR
66	TRANS
67	TRES
68	TR#E
69	TRI
70	ULTRA
71	VI
72	VICE

fig. 71. Ensemble des préfixes utilisés dans l'expérience

Remarques :

- Les préfixes et suffixes ont été, pour la plupart, extraits du "Bon Usage" (62)

- Si la liste des suffixes est relativement complète (mis à part les suffixes composés (cf. § IV.3. 3. 4.)) la liste des préfixes est au contraire relativement incomplète. Ceci est dû pour une part à l'aspect expérimental de l'étude, et, pour une autre part au fait que les préfixes conduisent souvent à une différence de sens notable.

1	-ATRE	===	ADJ	
2	ABLE	===	ADJ	
3	ACE	===	SUBS	
4	AC#E	===	ADJ	
5	ADE	===	SUBS	
6	AGE	===	SUBS	
7	AIE	===	SUBS	
8	AIL	===	SUBS	
9	AILLE	===	SUBS	
10	AILLER	===	VRB	
11	AIN	===	ADJ	SUBS
12	AINE	===	ADJ	SUBS
13	AIRE	===	ADJ	SUBS
14	AIS	===	ADJ	SUBS
15	AISON	===	SUBS	
16	AL	===	ADJ	
17	AN	===	ADJ	SUBS
18	ANCE	===	SUBS	
19	ANDE	===	SUBS	
20	ANDERIE	===	SUBS	
21	ANDIER	===	SUBS	
22	ANDI@ERE	===	SUBS	
23	ANE	===	ADJ	SUBS
24	ANT	===	ADJ	SUBS
25	ARD	===	ADJ	SUBS
26	AS	===	SUBS	
27	ASSE	===	SUBS	
28	ASSER	===	VRB	
29	AT	===	SUBS	
30	ATEUR	===	SUBS	
31	ATION	===	SUBS	
32	ATIQUE	===	ADJ	
33	ATRICE	===	SUBS	
34	ATURE	===	SUBS	
35	AUD	===	ADJ	SUBS
36	AUT	===	ADJ	SUBS
37	BILIT#E	===	SUBS	
38	BLE	===	ADJ	
39	BOND	===	ADJ	SUBS
40	BUS	===	SUBS	
41	CIR	===	VRB	
42	CULE	===	SUBS	
43	E	===	SUBS	
44	EAU	===	SUBS	
45	EL	===	ADJ	
46	ELER	===	VRB	
47	ELET	===	SUBS	
48	ELETTE	===	SUBS	
49	ELLE	===	SUBS	
50	EMENT	===	ADV	SUBS
51	ENCE	===	SUBS	
52	ER	===	SUBS	VRB
53	ERAIE	===	SUBS	
54	EREAU	===	SUBS	
55	ERELLE	===	SUBS	
56	ERESSE	===	SUBS	
57	ERET	===	SUBS	
58	ERIE	===	SUBS	
59	EROLE	===	SUBS	
60	ERON	===	SUBS	
61	ESQUE	===	ADJ	
62	ESSE	===	SUBS	
63	ET	===	SUBS	
64	ETER	===	VRB	
65	ET#E	===	SUBS	
66	ETTE	===	SUBS	
67	EUL	===	SUBS	
68	EUR	===	SUBS	
69	EUSE	===	ADJ	SUBS
70	EUX	===	ADJ	SUBS
71	@ERE	===	SUBS	
72	#E	===	ADJ	SUBS
73	#EE	===	SUBS	
74	#EEN	===	ADJ	SUBS
75	IBLE	===	ADJ	
76	ICHON	===	ADJ	SUBS
77	ICULE	===	SUBS	
78	IE	===	SUBS	
79	IEN	===	ADJ	SUBS
80	IENNE	===	ADJ	SUBS
81	IER	===	SUBS	
82	I@ERE	===	SUBS	
83	I#E	===	ADJ	SUBS
84	IF	===	ADJ	
85	IFIER	===	VRB	
86	IFLER	===	VRB	
87	IL	===	SUBS	
88	ILLE	===	SUBS	
89	ILLER	===	VRB	
90	ILLON	===	SUBS	
91	IN	===	ADJ	SUBS
92	INE	===	ADJ	SUBS
93	INER	===	VRB	
94	IOLE	===	SUBS	
95	IQUE	===	ADJ	
96	IR	===	VRB	
97	IS	===	SUBS	
98	ISE	===	SUBS	
99	ISER	===	VRB	
100	ISME	===	SUBS	
101	ISON	===	SUBS	
102	ISSAGE	===	SUBS	
103	ISSEMENT	===	SUBS	
104	ISTE	===	SUBS	
105	ITE	===	SUBS	
106	IT#E	===	SUBS	
107	ITION	===	SUBS	
108	ITUDE	===	SUBS	
109	IVE	===	ADJ	
110	MENT	===	ADV	SUBS
111	NICHER	===	VRB	
112	OCHER	===	VRB	
113	OIR	===	SUBS	
114	OIRE	===	SUBS	
115	OIS	===	ADJ	
116	OLE	===	SUBS	
117	ON	===	SUBS	
118	ONNER	===	VRB	
119	OSE	===	SUBS	
120	OT	===	ADJ	SUBS
121	OTE	===	ADJ	SUBS
122	OTER	===	VRB	
123	OTTE	===	ADJ	SUBS
124	OTTER	===	VRB	
125	OUILLER	===	VRB	
126	OYER	===	VRB	
127	RE	===	VRB	
128	SION	===	SUBS	
129	T#E	===	SUBS	
130	TION	===	SUBS	
131	TOIRE	===	ADJ	
132	U	===	ADJ	SUBS
133	UEUX	===	ADJ	
134	ULE	===	SUBS	
135	URE	===	SUBS	

fig. 72. Ensemble des suffixes utilisés dans l'expérience.

IV. 2. Choix du découpage d'un mot en radical + suffixe :

Ce problème est bien connu. Il est inhérent à l'analyse morphologique des langues naturelles basée sur le découpage des mots en "préfixe + radical + suffixe (5, 44, 68, 109) .

En effet, le problème auquel on est en général confronté est le choix d'un découpage d'un mot en "radical + suffixe", parmi plusieurs possibles. Par exemple, si "E" et "ITE" sont des suffixes retenus, sans autre contrainte, "VERITE" admettrait alors deux décompositions :

VERIT + E
VER + ITE

Afin de tenir compte des diverses altérations morphologiques auxquelles peuvent être soumis les radicaux et/ou suffixes lors de la dérivation et dans le but d'obtenir des familles de mots pertinentes aussi complètes que possibles, on est amené à s'accommoder de toutes les décompositions possibles des mots (fig. 73).

En effet, si l'on veut obtenir les familles de mots primitives suivantes :

$$F\ (ACH) = <\ ACH + AT,\ ACH + ETER,\ ACH + ETEUR\ >$$

$$F\ (CHRONOLOG) = <\ CHRONOLOG + IE,\ CHRONOLOG + IQUE,\ CHRONOLOG + IQUEMENT\ >$$

on constate qu'il y a lieu de recenser les chaînes suffixales suivantes :

AT, ETER, ETEUR, IE, IQUE, IQUEMENT

Or, parmi celles-ci, seules les chaînes suffixales suivantes :

AT, ETER, IE, IQUE,

appartiennent à l'ensemble **S** des suffixes répertoriés ;

ETEUR, IQUEMENT

étant des surcompositions de suffixes (cf. § IV. 3. 3. 4.)

	Mot	V(mot)	Radical	Suffixe	V(suffixe)	
121	ABJURER	VRB	ABJUR	ER	SUBS	VRB
122	ABJURER	VRB	ABJURER			
123	ABLATION	SUBS	ABL	ATION	SUBS	
124	ABLETTE	SUBS	ABL	ETTE	SUBS	
125	ABLATION	SUBS	ABLA	TION	SUBS	
126	ABLATION	SUBS	ABLATI	ON	SUBS	
127	ABLATION	SUBS	ABLATION			
128	ABLETTE	SUBS	ABLETT	E	SUBS	
129	ABLETTE	SUBS	ABLETTE			
130	ABLUTION	SUBS	ABLU	TION	SUBS	
131	ABLUTION	SUBS	ABLUTI	ON	SUBS	
132	ABLUTION	SUBS	ABLUTION			
133	ABN=EGATION	SUBS	ABN=EG	ATION	SUBS	
134	ABN=EGATION	SUBS	ABN=EGA	TION	SUBS	
135	ABN=EGATION	SUBS	ABN=EGATI	ON	SUBS	
136	ABN=EGATION	SUBS	ABN=EGATION			
137	ABOI	SUBS	ABOI			
138	ABOIEMENT	SUBS	ABOI	EMENT	ADV	SUBS
139	ABOIEMENT	SUBS	ABOIE	MENT	ADV	SUBS
140	ABOIEMENT	SUBS	ABOIEMENT			
141	ABOLIR	VRB	ABOL	IR	VRB	
142	ABOLITION	SUBS	ABOL	ITION	SUBS	
143	ABOLITION	SUBS	ABOLI	TION	SUBS	
144	ABOLIR	VRB	ABOLIR			
145	ABOLITION	SUBS	ABOLITI	ON	SUBS	
146	ABOLITION	SUBS	ABOLITION			
147	ABOMINABLE	ADJ	ABOMIN	ABLE	ADJ	
148	ABOMINABLE	ADJ	ABOMINA	BLE	ADJ	
149	ABOMINABLE	ADJ	ABOMINABLE			
150	ABONDANCE	SUBS	ABOND	ANCE	SUBS	
151	ABONDANT	ADJ	ABOND	ANT	ADJ	SUBS
152	ABONDER	VRB	ABOND	ER	SUBS	VRB
153	ABONDANCE	SUBS	ABONDANC	E	SUBS	
154	ABONDANCE	SUBS	ABONDANCE			
155	ABONDANT	ADJ	ABONDANT			
156	ABONDER	VRB	ABONDER			
157	ABONNEMENT	SUBS	ABONN	EMENT	ADV	SUBS
158	ABONNER	VRB	ABONN	ER	SUBS	VRB
159	ABONNEMENT	SUBS	ABONNE	MENT	ADV	SUBS
160	ABONNEMENT	SUBS	ABONNEMENT			
161	ABONNER	VRB	ABONNER			
162	ABORD	SUBS	ABORD			
163	ABORDER	VRB	ABORD	ER	SUBS	VRB
164	ABORDER	VRB	ABORDER			
165	ABOUCHER	VRB	ABOUCH	ER	SUBS	VRB
166	ABOUCHER	VRB	ABOUCHER			
167	ABOUTIR	VRB	ABOUT	IR	VRB	
168	ABOUTISSEMENT	SUBS	ABOUT	ISSEMENT	SUBS	
169	ABOUTIR	VRB	ABOUTIR			
170	ABOUTISSEMENT	SUBS	ABOUTISS	EMENT	ADV	SUBS
171	ABOUTISSEMENT	SUBS	ABOUTISSE	MENT	ADV	SUBS
172	ABOUTISSEMENT	SUBS	ABOUTISSEMENT			
173	ABOYER	VRB	ABOY	ER	SUBS	VRB
174	ABOYER	VRB	ABOYER			
175	ABRACADABRANT	ADJ	ABRACADABR	ANT	ADJ	SUBS
176	ABRACADABRANT	ADJ	ABRACADABRANT			
177	ABRASIF	ADJ	ABRAS	IF	ADJ	
178	ABRASIF	ADJ	ABRASIF			
179	ABREUVER	VRB	ABREUV	ER	SUBS	VRB
180	ABREUVOIR	SUBS	ABREUV	OIR	SUBS	

fig. 73. Décomposition multiple des mots en radical + suffixe triés selon le champ "radical".

On ne peut évidemment entreprendre de définir, sur l'ensemble de la langue, l'ensemble des chaînes suffixales permettant d'aboutir à des familles de mots complètes.

On peut cependant remarquer qu'une solution à ce problème réside précisément dans la conservation de toutes les décompositions des mots en radical + suffixe.

Si l'on reprend les exemples précédents, la famille F (ACH) se décompose alors comme suit :

$$F\ (ACH) = \langle\ ACH + AT,\ ACH + ETER\ \rangle \quad avec\ AT,\ ETER \in S$$
$$F\ (ACHET) = \langle\ ACHET + ER,\ ACHET + EUR\ \rangle \quad avec\ ER,\ EUR \in S$$

de même :

$$F\ (CHRONOLOG) = \langle\ CHRONOLOG + IE,\ CHRONOLOG + IQUE\ \rangle$$
$$F\ (CHRONOLOGIQU) = \langle\ CHRONOLOGIQU + E,\ CHRONOLOGIQU + EMENT\ \rangle$$

Dans tous ces cas, l'obtention de la famille la plus complète peut être assurée par réunion transitive. Cette opération ne peut cependant être effectuée sans quelques précautions que nous verrons dans la suite de l'exposé.

Nous donnons ci-après (fig. 74) un extrait de quelques familles primitives suffixales. Sur la figure on distingue successivement :
- le numéro d'ordre attribué à la famille ;
- le radical autour duquel celle-ci s'est formée ;
- les éléments qui la constituent enfin.

IV. 3. Obtention des familles de mots propres :

Partant des familles de mots primitives, notre démarche va désormais consister à engendrer, à partir de celles-ci, des familles de mots "propres" (c'est à dire correctes) auxquelles nous voulons précisément aboutir.

En effet, l'observation des familles primitives nous montre que l'on est en présence de deux situations différentes (qui peuvent par ailleurs se combiner) appelant chacune un type de traitement différent. On trouve en effet, outre les familles primitives qui correspondent à des familles propres et complètes (et qui ne demandent par conséquent aucun traitement particulier) :

2	¬ACRE	*	¬ACRE	¬ACRET#E		
3	¬AG	*	¬AGE	¬AG#E		
4	¬AN	*	¬ANE ¬ANON	¬ANERIE	¬ANESSE	¬ANIER
5	¬APRE	*	¬APRE	¬APRET#E		
6	¬IL	*	¬ILE	¬ILOT		
7	A	*	ABUS	AIL	AINE	AISE
8	A¬IN	*	A¬INESSE	A¬IN#E		
9	AB	*	ABAT	ABONNER	ABOYER	
10	AB¬IM	*	AB¬IME	AB¬IMER		
11	ABAISS	*	ABAISSEMENT	ABAISSER		
12	ABAND	*	ABANDON	ABANDONNER		
13	ABATT	*	ABATTAGE	ABATTEMENT	ABATTOIR	ABATTRE
14	ABB	*	ABBESSE	ABB#E		
15	ABL	*	ABLATION	ABLETTE		
16	ABOI	*	ABOI	ABOIEMENT		
17	ABOL	*	ABOLIR	ABOLITION		
18	ABOND	*	ABONDANCE	ABONDANT	ABONDER	
19	ABONN	*	ABONNEMENT	ABONNER		
20	ABORD	*	ABORD	ABORDER		
21	ABOUT	*	ABOUTIR	ABOUTISSEMENT		
22	ABREUV	*	ABREUVER	ABREUVOIR		
23	ABR#EG	*	ABR#EGER	ABR#EG#E		
24	ABROG	*	ABROGATION	ABROGER		
25	ABSENT	*	ABSENT	ABSENTER		
26	ABSOLU	*	ABSOLU	ABSOLUMENT	ABSOLUTION	
27	ABSORB	*	ABSORBANT	ABSORBER		
28	ABSTEN	*	ABSTENIR	ABSTENTION		
29	ABUS	*	ABUS	ABUSER	ABUSIF	
446	BIGARR	*	BIGARREAU	BIGARR#E	BIGARRURE	
447	BIJOUT	*	BIJOUTERIE	BIJOUTIER		
448	BIL	*	BILAN	BILE	BILER	
449	BILL	*	BILL BILLON	BILLARD BILLOT	BILLE	BILLET
450	BIN	*	BINAIRE	BINETTE		
451	BIOLOG	*	BIOLOGIE	BIOLOGIQUE		
452	BIS	*	BISE	BISEAU	BISON	
453	BISCUIT	*	BISCUIT	BISCUITERIE		

fig. 74. Extrait des familles primitives suffixales.

a) des familles primitives qui demandent à être scindées en vue d'obtenir des familles propres :

Exemple :

$$F\ (SAV) = \langle\ SAVANT,\ SAVEUR,\ SAVOIR,\ SAVON\ \rangle$$

b) des familles primitives qui demandent à être réunies afin d'obtenir des familles complètes :

Exemple :

$$F\ (CONFECTI) = \langle\ CONFECTION,\ CONFECTIONNER\ \rangle$$
$$F\ (CONFECTIONN) = \langle\ CONFECTIONNER,\ CONFECTIONNEUR\ \rangle$$

Notons que les traitements associés au point b) ne peuvent être entrepris qu'après avoir effectué les traitements relatifs au point a). En effet, l'option contraire nous amènerait à effectuer un travail supplémentaire puisque nous aurions à disjoindre des familles qui l'étaient déjà.

Exemple :

$$F\ (ACCESS) = \langle\ ACCESSIBLE,\ ACCESSOIRE\ \rangle$$
$$F\ (ACCESSOIR) = \langle\ ACCESSOIRE,\ ACCESSOIREMENT\ \rangle$$

IV. 3. 1. Notion de schéma dérivationnel :

Soit la famille de mots :

$F(r) = \langle\ m_1,\ m_2,\ \ldots,\ m_n\ \rangle$

On peut écrire :

$F(r) = \langle\ r + s_1,\ r + s_2,\ \ldots.,\ r + s_n\ \rangle$, $s_i \in S$

On appelle schéma dérivationnel l'ensemble des suffixes :

$D(r) = \langle\ s_1,\ s_2,\ \ldots,\ s_n\ \rangle$

Si la famille F(r) est primitive, D(r) sera dit schéma dérivationnel primitif.

Un même schéma dérivationnel peut être associé à plusieurs radicaux différents. Nous noterons alors indifféremment D ou D(r) le schéma dérivationnel, selon que l'on a en vue de spécifier ou non le radical auquel il est associé.

Exemples :

$$\begin{cases} D\ (ELECTOR) = < AL,\ AT > \\ D\ (EPISCOP) = < AL,\ AT > \end{cases}$$

$$\begin{cases} D\ (ARPENT) = < b,\ AGE,\ ER,\ EUR > \\ D\ (LABOUR) = < b,\ AGE,\ ER,\ EUR > \end{cases}$$

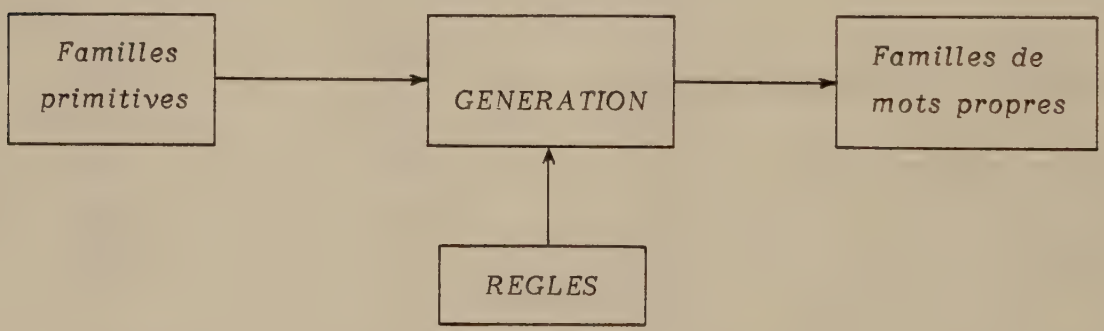
$$D\ (AVO) = < INE,\ IR >$$
$$D\ (BAS) = < b,\ CULE,\ E,\ ER,\ TION >$$

La figure 75 offre des exemples de schémas dérivationnels ainsi obtenus.

La notion de schéma dérivationnel est, ainsi que nous allons le voir, à la base des heuristiques que nous allons présenter.

IV. 3. 2. Explosion des familles de mots primitives :

Nous nous proposons d'établir un algorithme qui nous permette d'engendrer des familles (ou sous-familles) propres de mots à partir des familles primitives



Par exemple, à :

$$F\ (INTEGR) = < INTEGRAL,\ INTEGRALEMENT,\ INTEGRATION,\ INTEGRER,\ INTEGRITE >$$

on veut faire correspondre :

$$F_1\ (INTEGR) = < INTEGRAL,\ INTEGRALEMENT >$$
$$F_2\ (INTEGR) = < INTEGRAL,\ INTEGRATION,\ INTEGRER >$$
$$F_3\ (INTEGR) = < INTEGRITE >$$

997	#EDUC	* ATEUR	ATIF	ATION		
998	RECTIFIC	* ATEUR	ATIF	ATION		
999	COOP#ER	* ATEUR	ATIF	ATION	ATIVE	ER
1000	L#EGISL	* ATEUR	ATIF	ATION	ATRICE	ATURE
1001	INDIC	* ATEUR	ATIF	ATION	ATRICE	E
1002	ADMINISTR	* ATEUR * #E	ATIF	ATION	ATRICE	ER
1003	SP#ECUL	* ATEUR	ATIF	ATION	ER	
1004	ALTERN	* ATEUR	ATIF	ATIVE	ER	
1005	D#ENONCI	* ATEUR	ATION			
1006	IDENTIFIC	* ATEUR	ATION			
1007	M#EDI	* ATEUR	ATION			
1008	R#EGUL	* ATEUR	ATION			
1009	TARIFIC	* ATEUR	ATION			
1010	V#ERIFIC	* ATEUR	ATION			
1011	CONSERV	* ATEUR * ER	ATION ERIE	ATOIRE	ATRICE	E
1012	OBSERV	* ATEUR	ATION	ATOIRE	ATRICE	ER
1013	OP#ER	* ATEUR	ATION	ATOIRE	ATRICE	ER
1014	PR#EPAR	* ATEUR	ATION	ATOIRE	ATRICE	ER
1015	#EL#EV	* ATEUR	ATION	ATRICE		
1016	G#EN#ER	* ATEUR	ATION	ATRICE		
1017	NOV	* ATEUR	ATION	ATRICE		
1018	PERTURB	* ATEUR	ATION	ATRICE		
1019	PR#EDIC	* ATEUR	ATION	ATRICE		
1020	CR#E	* ATEUR	ATION	ATRICE	ATURE	ER
1021	EXPORT	* ATEUR	ATION	ATRICE	ER	
1022	IMPORT	* ATEUR	ATION	ATRICE	ER	
1023	R#ENOV	* ATEUR	ATION	ATRICE	ER	
1024	USURP	* ATEUR	ATION	ATRICE	ER	
1025	ACCUS	* ATEUR	ATION	ATRICE	ER	#E

fig. 75. Extrait de quelques schémas dérivationnels.

En utilisant les schémas dérivationnels, on peut écrire :

$$D\ (INTEGR) = <\ AL,\ ALEMENT,\ ATION,\ ER,\ ITE\ >$$

et

$$D_1\ (INTEGR) = <\ AL,\ ALEMENT\ >$$
$$D_2\ (INTEGR) = <\ AL,\ ATION,\ ER\ >$$
$$D_3\ (INTEGR) = <\ ITE\ >$$

Remarque : On peut noter l'absence dans cet exemple des mots INTEGRALE et INTEGRALITE. La raison en est qu'ils admettent respectivement pour suffixes ALE et ALITE, suffixes qui n'ont pas été répertoriés comme tels à ce stade du traitement, et qui correspondent en fait à des suffixes composés. Ce point a été effleuré dans le paragraphe IV. 2. et sera développé au paragraphe IV. 3. 3. 4.

IV. 3. 2. 1. Schéma dérivationnel impropre :

Un schéma dérivationnel $D = < s_1, s_2, \ldots, s_n >$ est dit impropre s'il n'existe aucun radical r tel que D(r) corresponde à une famille de mots propre.

Ainsi le schéma dérivationnel de l'exemple précédent :

$$D = <\ AL,\ ALEMENT,\ ATION,\ ER,\ ITE\ >$$

est impropre.

On peut étendre cette notion à des sous-schémas dérivationnels. Considérons par exemple le sous-schéma dérivationnel

$$D = <\ ATION,\ ETTE\ >$$

On peut alors trouver :

$$F\ (ABL) = <\ ABLATION,\ ABLETTE\ >$$
$$F\ (CASS) = <\ CASSATION,\ CASSETTE,\ \ldots\ >$$
$$F\ (OPER) = <\ OPERATION,\ OPERETTE,\ \ldots\ >$$
$$F\ (VIOL) = <\ VIOLATION,\ VIOLETTE,\ \ldots\ >$$

mais aucun de ces radicaux ne donne naissance à une famille ou sous-famille de mots correcte. Le schéma ou sous-schéma D = <ATION, ETTE> est donc impropre et par suite impossible.

Ceci nous amène à émettre l'hypothèse fondamentale suivante :

Hypothèse I :

La cooccurrence, au sein d'un même schéma dérivationnel correct, de certains suffixes est impossible.

Si nous reprenons les exemples précédents :

<AL, ER, ITE> , <ATION, ETTE> , etc..

sont impossibles.

Nous dirons que les deux suffixes ATION et ETTE sont généralement incompatibles. De même le triplet (AL, ER, ITE) est généralement incompatible.

L'obtention des familles de mots propres, à partir des familles primitives, peut dès lors être assurée par un calcul sur les schémas dérivationnels primitifs sans tenir compte des radicaux.

Il subsistera cependant des cas marginaux qui ne pourront être traités que "manuellement".

Ces hypothèses vont nous permettre, ainsi que nous allons le voir, de créer les règles de production des familles de mots, à partir des familles primitives.

IV. 3. 2. 2. Matrice de corrélation binaire des suffixes :

Afin de dégager l'ensemble des doublets incompatibles d'une part, et de créer par là même les règles de production des familles de mots, on établit une matrice qui fait le bilan des corrélations binaires relatives aux suffixes des schémas primitifs.

Celle-ci se présente sous la forme d'un tableau booléen M_S correspondant au produit **S** x **S** : à chaque suffixe $s \in \mathbf{S}$ correspond dans M_S une ligne et une colonne.

Cette matrice est ensuite remplie comme suit : si deux suffixes s_i et s_j apparaissent simultanément dans au moins un schéma primitif, alors

$M_S(s_i, s_j) = 1$; sinon $M_S(s_i, s_j) = 0$.

Nous donnons figure 76 un extrait de la matrice de corrélation des suffixes obtenue à partir des schémas dérivationnels primitifs. Le listage fait apparaître, dans l'ordre et de gauche à droite, le nom de la ligne (ou colonne), les colonnes (ou lignes), le nombre d'occurrences des couples de suffixes et enfin les radicaux associés.

La matrice de corrélation binaire sera dite locale si elle est associée à un seul schéma dérivationnel.

Notons enfin que la matrice de corrélation est symétrique par rapport à la diagonale principale. Seule une partie est nécessaire, mais pour des commodités techniques la matrice entière est utilisée.

IV. 3. 2. 3. Matrice d'incompatibilité des suffixes ; Matrice de compatibilité des suffixes :

L'examen de la matrice de corrélation des suffixes M_S fait apparaître trois situations différentes :

a) certains points de M_S sont généralement incompatibles ;
Exemples :

(ATION, ETTE), (ASSER, ENCE), ...

b) certains points sont généralement compatibles ;
Exemples :

(ISER, ITE), (AILLE, ALLER), ...

c) enfin, une dernière catégorie de points de la matrice M_S sont indéterminés: ils sont en effet compatibles pour une partie des radicaux et incompatibles pour l'autre. Nous dirons que ces points sont ambigus.

Exemples :

(ATION, ER), (ERE, IQUE), ...

Les points compatibles et incompatibles sont respectivement consignés dans les matrices de compatibilité et d'incompatibilité. Celles-ci ont évidemment la même structure que M_S. Nous les noterons respectivement M_C et M_I (cf. fig. 77 et 78).

32	ATIQUE	***		2	*	FAN	HIER			
			AGE	1	*	SCI				
			AISON	1	*	LUN				
			ANT	1	*	ERR				
			AT	1	*	CLIM				
			ATION	1	*	INFORM				
			E	8	*	AROM	DOGM	DRAM	ERR	#ENIGM
					*	LUN	LYMPH	SCI		
			EMENT	1	*	ERR				
			ENCE	1	*	SCI				
			ER	4	*	ERR	FAN	INFORM	SCI	
			ERIE	1	*	SCI				
			ETTE	1	*	LUN				
			EUR	2	*	ERR	SCI			
			EUX	1	*	AQU				
			#E	1	*	LUN				
			ITE	1	*	MORGAN				
			ON	1	*	FAN				
			URE	1	*	SCI				
33	ATRICE	***		2	*	DESSIN	TEST			
			ABLE	3	*	ADMIR	CONSOMM	UTILIS		
			AIRE	1	*	SECT				
			AL	2	*	G#EN#ER	TRIOMPH			
			ANCE	2	*	CR#E	IMPORT			
			ANT	4	*	DOMIN	IMPORT	PROVOC	TRIOMPH	
			ASSER	1	*	D#EL				
			ATEUR	32	*	ACCUMUL	ACCUS	ADMINISTR	ADMIR	CONSERV
					*	CONSOMM	CR#E	DESSIN	D#EL	DOMIN
					*	EXAMIN	EXPORT	#EL#EV	G#EN#ER	IMPORT
					*	INDIC	L#EGISL	NOV	OBSERV	OP#ER
					*	PERTURB	PR#EDIC	PR#EPAR	PROVOC	R#ENOV
					*	SECT	SPECT	TEST	T#EL#ESPECT	TRIOMPH
					*	USURP	UTILIS			
			ATION	23	*	ACCUS	ADMINISTR	ADMIR	CONSERV	CONSOMM
					*	CR#E	DOMIN	EXPORT	#EL#EV	G#EN#ER
					*	IMPORT	INDIC	L#EGISL	NOV	OBSERV
					*	OP#ER	PERTURB	PR#EDIC	PR#EPAR	PROVOC
					*	R#ENOV	USURP	UTILIS		
			ATURE	2	*	CR#E	L#EGISL			
			E	4	*	CONSERV	INDIC	SECT	TRIOMPH	
			ER	20	*	ACCUMUL	ACCUS	ADMINISTR	ADMIR	CONSERV
					*	CONSOMM	CR#E	DESSIN	DOMIN	EXAMIN
					*	EXPORT	IMPORT	OBSERV	OP#ER	PR#EPAR
					*	R#ENOV	TEST	TRIOMPH	USURP	UTILIS
			ERIE	1	*	CONSERV				
			ETTE	1	*	OP#ER				
			EUR	1	*	SECT				
			EUX	1	*	G#EN#ER				
			#E	3	*	ACCUS	ADMINISTR	CONSOMM		
			IBLE	1	*	INDIC				
			I#E	1	*	D#EL				
			INE	1	*	CANT				
			OLE	1	*	CR#E				
			ON	1	*	CANT				
			ONNER	1	*	CANT				
			TION	1	*	PR#EDIC				
98	ISE	***		9	*	BALOURD	EXPERT	FRIAND	GOURMAND	MARCHAND
					*	MER	MIGNARD	VAL	VANTARD	
			ABLE	1	*	VAL				
			AGE	1	*	MARQU				
			ANT	1	*	MARQU				
			AUT	1	*	ASS				
			E	9	*	B-ET	BALOURD	BANQU	MA-ITR	MARQU
					*	PR-ETR	SOTT	TRA-ITR	VANTARD	
			ENCE	1	*	VAL				
			ER	5	*	CONVOIT	GOURMAND	HANT	MARCHAND	MARQU
			ESSE	3	*	MA-ITR	PR-ETR	TRA-ITR		
			ET	2	*	BANQU	VAL			
			ETER	1	*	BANQU				
			ETTE	1	*	BANQU				
			EUR	1	*	VAL				
			IE	1	*	BRA				
			IER	1	*	BANQU				
			ILLER	1	*	BRA				
			IN	2	*	CHEM	DEV			
			INER	2	*	CHEM	DEV			
			IR	1	*	FRANCH				
			IS	18	*	ANGLA	DEV	#ECOSSA	FRA	FRAN$CA
					*	HONGRO	INCOMPR	IND#EC	JAPONA	LYONNA
					*	MALAPPR	MARQU	M#EPR	POLONA	PORTUGA
					*	PUNA	REPR	SIAMO		
			ISER	5	*	BRA	DEV	FRA	M#EPR	REPR
			ISSEMENT	1	*	FRANCH				
			OCHER	1	*	REPR				
			OIR	1	*	DEV				

fig. 76. Matrice de corrélation des suffixes

136		*** -ATRE	4	* BLEU	FOL	GRIS	NOIR	
		AISON	1	* D#ECLIN				
		ANDIƏERE	1	* FIL				
		ATURE	1	* FIL				
		AUD	5	* COURT	FIN	LOURD	NOIR	ROUGE
		CIR	3	* DUR	NOIR	OBSCUR		
		ELET	2	* DRAP	GANT			
		EPESSE	2	* DEVIN	FORT			
		ESQUE	2	* ROMAN	SOLDAT			
		ET#E	6	* DUR	IMPUR	PUR	S-UR	SAINT
				* SOUVERAIN				
		IVE	1	* SPORT				
1	-ATRE	***	4	* BLEU	FOL	GRIS	NOIR	
		AILLE	1	* GRIS				
		AUD	1	* NOIR				
		CIR	1	* NOIR				
		ETTE	2	* FOL	GRIS			
		EUR	2	* BLANCH	VERD			
		IE	1	* FOL				
		IER	2	* OLIV	VERD			
		INE	1	* OLIV				
		IGUE	1	* FOL				
		IR	3	* BLANCH	JAUN	VERD		
		ISSAGE	1	* BLANCH				
		ISSEMENT	1	* BLANCH				
		ONNER	1	* GRIS				
		OYER	1	* VERD				
17	AN	*** ANE	7	* ANGLIC	MAHOM#ET	M#EDI	MUSULM	ORG
				* PARTIS	SULT			
18	ANCE	*** ANT	44	* ABOND	AMBI	AMBUL	ASSIST	ATTEN
				* BIENFAIS	BIENVEILL	CONCORD	CONTEN	CORRESPOND
				* CROISS	DESCEND	D#EFAILL	DIST	ENF
				* EXTRAVAG	#ECH#E	#EL#EG	G#ER	IGNOR
				* IMPORT	IMPUISS	IND#EPEND	INST	INSUFFIS
				* INTEND	LUXURI	M#EFI	MOUV	OB#EISS
				* PART	PERSIST	PLAIS	PR#EPOND#ER	PR#EVOY
				* PUISS	RECONNAISS	RENAISS	R#ESIST	SOUFFR
				* SUBSIST	TEND	VAC	VIGIL	
		EUR	5	* ASSUR	CONTEN	JOUISS	R#EASSUR	TEND
		EUSE	1	* JOUISS				
		#E	2	* AIS	D#ESESP#ER			
		IBLE	1	* NUIS				
		I#E	1	* PIT				
		U	1	* CONTEN				
19	ANDE	*** ANDIER	1	* CONTREB				
		ATEUR	1	* MULTIPLIC				
		ATION	2	* MULTIPLIC	PROPAG			
		IR	1	* OFFR				
		IT#E	1	* MULTIPLIC				
21	ANDIER	*** ANDE	1	* CONTREB				
		ANDIƏERE	1	* VIV				
		ER	1	* TAILL				
22	ANDIƏERE	***	1	* FIL				
		ANDIER	1	* VIV				
		ATEUR	1	* FIL				
		ATURE	1	* FIL				
		ER	1	* FIL				
30	ATEUR	*** ANDE	1	* MULTIPLIC				
		ANDIƏERE	1	* FIL				
		AT	1	* S#EN				
		ATION	75	* ACCUS	ADMINISTR	ADMIR	ADOR	AGIT
				* ANIM	ASPIR	ATT#ENU	AVI	COLLABOR
				* CONCILI	CONDENS	CONSERV	CONSOMM	COOP#ER
				* CR#E	D#ECOR	D#ENATUR	D#ENOMIN	D#ENONCI
				* DISTILL	DOMIN	DON	EXP#ERIMENT	EXPORT

fig. 77. Matrice de compatibilité des suffixes.

1	-ATRE	***	ER	-1 * GRIS				
			OT	-1 * MUL				
2	ABLE	***	ADE	-1 * PASS				
			AIRE	-5 * AFF	ANNUL	HONOR	NOT	VULN#ER
			ANDE	-1 * LAV				
			ASSE	-1 * POT				
			ASSER	-1 * POT				
			CIR	-1 * DUR				
			EAU	-1 * POT				
			EL	-1 * FR#EJUDICI				
			EREAU	-1 * PASS				
			ET	-3 * LIVR	MIN	VAL		
			ET#E	-1 * DUR				
			ETTE	-5 * CUR	LAV	MIN	MOUILL	SERVI
			ƏERE	-1 * MANI				
			IBLE	-1 * PASS				
			IE	-1 * INCUR				
			IER	-4 * OUVR	PORT	POT	RENT	
			IƏERE	-4 * MIN	OUVR	PORT	RENT	
			IF	-2 * PASS	RESPECT			
			IFIER	-1 * NOT				
			ILLE	-1 * FLOTT				
			ILLER	-1 * INST				
			ILLON	-2 * OUR	PORT			
			IN	-1 * POT				
42	CULE	***		-2 * BAS	PARTI			
			AL	-1 * PARTI				
			E	-2 * BAS	PARTI			
			EL	-1 * PARTI				
			ER	-1 * BAS				
			TION	-3 * BAS	PARTI	TENTA		
43	E	***	ACE	-3 * LIM	PAL	ROS		
			AIE	-2 * PAG	SAG			
			AILLER	-3 * JOU	MITR	RIM		
			AINE	-6 * AUB	BOURD	FONT	GERM	MIT
				* MOR				
			AN	-7 * ANT	BIL	CADR	MERL	ORG
				* PAL	TART			
			ANDIER	-1 * TAILL				
			ANE	-3 * MEMBR	ORG	SOUT		
			AUD	-3 * FIN	MAR	PAT		
			CULE	-2 * BAS	PARTI			
			ELET	-3 * CHAP	ORG	PIP		
			ELETTE	-2 * C-OT	PIP			
			ENCE	-6 * D#ECAD	LIC	PRUD	SCI	SENT
				* VIOL				
			ERELLE	-1 * SAUT				
			EROLE	-2 * BAND	CASS			
			EUL	-1 * FILL				
			IBLE	-3 * IMPASS	INDIC	TERR		
136		***	ACE	-2 * FOU	PAL			
			AC#E	-1 * VIOL				
			AIL	-4 * COR	#EVENT	PORT	SOUPIR	
			AIS	-3 * FRANC	PAL	PAN		
			AN	-4 * BOUC	CHOU	PAL	PAYS	
			ANE	-5 * BAN	BUT	CAB	CHIC	PLAT
			AS	-3 * CAB	FRAC	GALET		
			ATIQUE	-2 * FAN	HIER			
			BILIT#E	-1 * NOTA				
			CULE	-2 * BAS	PARTI			
			ENCE	-4 * CAR	POT	VAL	VIOL	
			EREAU	-1 * BORD				
			ERELLE	-1 * CHANT				
			ƏERE	-2 * ART	COL			
			IEN	-3 * MAINT	NORMAL	SAUR		
			IENNE	-2 * C#ESAR	NORMAL			
			IL	-5 * BAR	GR#ES	ORGUE	SOLE	SOMME
			ILLE	-9 * BAC	BANDER	BROU	CORNE	FEU
				* FOU	LENT	VAN	VIE	
			ILLON	-6 * AIGU	BROU	CAR	OUR	PORT
				* SOU				
			OLE	-4 * BANDER	COUP	GOND	ROUGE	
			OSE	-1 * COUPER				
			OT	-10 * BILL	CAB	CAGE	CAL	CAP
				* CHIC	CUL	PALET	PIC	ROS
			OTE	-2 * CAP	JUGE			
			OTER	-5 * CAB	CAP	CHIC	PIC	RAD
			OTTE	-4 * CAR	CUL	LIN	MARC	
			OUILLER	-2 * CHAT	GAZ			
			RE	-1 * FOND				
			ULE	-1 * ROT				

fig. 78. Matice d'incompatibilité des suffixes.

IV. 3. 2. 4. Production des familles de mots propres :

Rappelons sur un exemple le but que l'on se propose d'atteindre.
Partant de

$$D\ (CHIM) = <\ ERE,\ IE,\ IQUE,\ ISTE\ >$$

on veut obtenir :

$$D_1\ (CHIM) = <\ ERE\ >$$
$$D_2\ (CHIM) = <\ IE,\ IQUE,\ ISTE\ >$$

Plus généralement, partant d'un schéma dérivationnel primitif $D = < s_1, s_2, \ldots, s_n >$ on veut engendrer, si D est incorrect, plusieurs sous-schémas dérivationnels corrects $D_1, D_2, \ldots, D_k$ qui donneront des familles ou sous familles de mots propres.

Pour cela, nous opérons comme suit :

Nous associons à $D = < s_1, s_2, \ldots, s_n >$ une matrice de corrélation des suffixes M_D, dite locale de dimension n x n (n = card D). Notons que l'on aura :

$$M_D(s_i, s_j) = 1 \qquad \forall i, j \in (1, n)$$

Exemple :

Soit $D\ (CHIM) = <\ ERE,\ IE,\ IQUE,\ ISTE\ >$

La matrice associée sera :

M_D	ERE	IE	IQUE	ISTE
ERE	1	1	1	1
IE	1	1	1	1
IQUE	1	1	1	1
ISTE	1	1	1	1

Cet exemple servira d'illustration dans la suite de l'exposé.

La méthode d'obtention des familles correctes est basée sur une modification dynamique de M_D par les matrices de compatibilités M_C et M_I.

a) Modification de la matrice de corrélation locale M_D par la matrice des incompatibilités M_I :

Cette première modification de M_D consiste à éliminer de M_D les points qui sont incompatibles.

Chaque point (s_i, s_j) de la matrice M_D est comparé à son homologue (s_i', s_j') de la matrice M_I des incompatibilités. Si $M_I(s_i', s_j') = 0$ alors $M_D(s_i, s_j)$ prend la valeur zéro, sinon il reste inchangé. Nous noterons M_D^I la matrice M_D ainsi modifiée. Notons que M_D^I ne contient désormais que des points qui sont soit compatibles, soit ambigus.

Ainsi la matrice M_D de l'exemple précédent devient :

M_D^I	*ERE*	*IE*	*IQUE*	*ISTE*
ERE	1	0	1	0
IE	0	1	1	1
IQUE	1	1	1	1
ISTE	0	1	1	1

A partir du schéma dérivationnel et de la matrice M_D^I associée, on engendre tous les sous-schémas dérivationnels inclus dans le schéma initial et qui vérifient les relations d'incompatibilité de la matrice de corrélation locale M_D^I.

Parmi les sous-schémas engendrés, sont évidemment éliminés ceux qui sont strictement inclus dans un autre.

Ainsi, si nous reprenons notre exemple, on aura :

$D_1 = \langle ERE, IQUE \rangle$
$D_2 = \langle IE, IQUE, ISTE \rangle$

Nous sommes donc, à ce stade, passés de la famille primitive :

$$F\ (CHIM) = <\ CHIMERE,\ CHIMIE,\ CHIMIQUE,\ CHIMISTE\ >$$

aux familles :

$$F_1(CHIM) = <\ CHIMERE,\ CHIMIQUE\ >$$
$$F_2(CHIM) = <\ CHIMIE,\ CHIMIQUE,\ CHIMISTE\ >$$

avec cependant F_1 incorrecte et F_2 correcte. Nous sommes donc passés de la sorte de schémas dérivationnels primitifs où les suffixes étaient deux à deux compatibles, incompatibles ou indéterminés à des schémas dérivationnels intermédiaires où les suffixes sont deux à deux compatibles ou indéterminés seulement.

Notons enfin que nous effectuons là un recouvrement et non une partition d'un schéma dérivationnel, de sorte que les schémas retenus ne sont pas forcément disjoints.

b) Modification de la matrice M_D^I par la matrice des compatibilités M_C :

Pour éviter la production de familles de mots incorrectes (F_1 dans l'exemple traité) qui est due à l'existence dans la matrice M_D^I de points ambigus, on fait intervenir la matrice des compatibilités M_C afin de rendre ces derniers, selon le cas, totalement compatibles ou incompatibles.

Le problème qui se pose est donc la détection des points ambigus et leur transformation en points compatibles ou incompatibles.

La détection de ces points dans M_D^I est simple : tout point (s_i, s_j) tel que $M_D^I(s_i, s_j) = 0$ est ambigu, si son homologue (s_i', s_j') dans la matrice des compatibilités est tel que $M_C(s_i', s_j') = 0$. Ainsi, les points ambigus dans l'exemple traité sont : (ERE, IQUE) et (IQUE, ISTE).

Hypothèse II :

Notons :
R_C la relation de compatibilité entre deux suffixes,
R_I la relation d'incompatibilité.
Soient x, y et z trois suffixes appartenant au schéma dérivationnel D.

La levée de l'ambiguïté sur les points indéterminés de la matrice M_D^I repose sur l'hypothèse suivante :

Si x et y sont compatibles et si y et z sont incompatibles alors x et z sont incompatibles.

Cela peut s'écrire :

$$x\ R_C\ y \quad \text{et} \quad y\ R_I\ z \quad \Longrightarrow \quad x\ R_I\ z$$

Exemple :

Considérons les trois suffixes ON, ONNER et ATION.

On a, si l'on consulte les matrices de compatibilité des suffixes :

$$(ON)\ R_C\ (ONNER) \quad et \quad (ONNER)\ R_I\ (ATION)$$

On peut alors déduire la nature de la troisième relation (ON - ATION) et qui sera dans le cas présent une relation d'incompatibilité. Cela permet donc d'éliminer des occurrences telles que < COT + ON, COT + ATION > sans pour autant indire des occurrences telles que < AVI + ON, AVI + ATION > .

Remarque :

Nous pourrions imaginer que la relation R_C est transitive :

$$x\ R_C\ y \quad \text{et} \quad y\ R_C\ z \quad \Longrightarrow \quad x\ R_C\ z$$

cela ne peut cependant être retenu, ainsi que nous le verrons, en raison de la polysémie (cf. IV. 3. 3. 1.). On pourrait, en effet, imaginer une famille :

$$F = \langle r + x,\ r + y,\ r + z \rangle$$

où r + y est polysème. Une telle famille devrait alors se décomposer en :

$$F_1 = \langle r + x,\ r + y \rangle$$

$$F_2 = \langle r + y,\ r + z \rangle$$

Exemple :

$$F(ACTI) = \langle ACTI + ONNER,\ ACTI + ON,\ ACTI + ONNAIRE \rangle$$

De même l'hypothèse :

$$x\ R_I\ y \quad \text{et} \quad y\ R_I\ z \quad \Longrightarrow \quad x\ R_I\ z$$

ne peut être retenue, ainsi que nous pouvons le voir sur l'exemple suivant :

$$F(POST) = \langle POST + AL,\ POST + URE,\ POST + IERE \rangle$$

où l'on a :

$$(AL)\ R_I\ (URE) \quad et \quad (URE)\ R_I\ (IERE)$$

alors que AL est compatible avec IERE.

L'hypothèse II va nous permettre de modifier la matrice M_D^I en transformant les points ambigus en points soit compatibles, soit incompatibles. Il subsistera cependant, dans de rares cas, des ambiguïtés qui ne pourront être levées que "manuellement".

La matrice M_D^I modifiée de la sorte constituera la matrice de corrélation locale M_D^{IC} définitive associée au schéma dérivationnel D.

La mise en oeuvre de cette hypothèse se traduit comme suit :

soient x et y deux suffixes tels que $M_C(x, y) = 1$

on a alors : $\forall s \in S$

$$M_D^{IC}(s, x) = M_D^{IC}(s, y) = M_D^I(s, x) \wedge M_D^I(s, y)$$

Ainsi, nous levons l'ambiguïté sur le couple (s, x) si le couple (s, y) est non ambigu. Nous ne pouvons cependant pas aboutir à un résultat non-ambigu si les deux couples (s, x) et (s, y) sont ambigus. Aussi, par défaut de règle d'incompatibilité, un point ambigu sera considéré comme compatible.

Remarquons que par suite de la non-transitivité de la relation R_C de compatibilité des suffixes, nous aurons, si :

$$x, y, z \in S \qquad M_C(x, y) = 1$$
$$M_C(y, z) = 1$$

alors : $\forall s \in S$

$$M_D^{IC}(s, x) = M_D^I(s, x) \wedge M_D^I(s, y)$$

$$M_D^{IC}(s, y) = M_D^I(s, x) \wedge M_D^I(s, y) \wedge M_D^I(s, z)$$

$$M_D^{IC}(s, z) = M_D^I(s, z) \wedge M_D^I(s, y)$$

Cela nous amène à définir la matrice M_D^{IC} au moyen des relations récurrentes :

si $M_C(x, y) = 1$ alors $\forall s \in S$

$$\left.\begin{aligned} M_D^{IC}(s, x) &= M_D^{IC}(s, x) \wedge M_D^{I}(s, y) \\ M_D^{IC}(s, y) &= M_D^{IC}(s, y) \wedge M_D^{I}(s, x) \end{aligned}\right\} (1)$$

avec au départ :

$$M_D^{IC} = M_D^{I}$$

Reprenons la matrice M_D^I de l'exemple du paragraphe IV. 3. 2. 4.

Nous avons donc :

points compatibles : (IE, IQUE)
(IE, ISTE)

points ambigus : (ERE, IQUE)
(IQUE, ISTE)

La matrice M_D^{IC} est au départ identique à M_D^I.

L'application de (1) donne :

$$M_D^{IC}(ERE, IQUE) = M_D^{IC}(ERE, IQUE) \wedge M_D^{I}(ERE, IE) = 0$$

$$M_D^{IC}(IQUE, ISTE) = M_D^{IC}(IQUE, ISTE) \wedge M_D^{I}(IE, IQUE) \wedge M_D^{I}(IE, ISTE) = 1$$

La matrice de corrélation locale M_D^I devient donc :

M_D^{IC}	ERE	IE	IQUE	ISTE
ERE	1	0	0	0
IE	0	1	1	1
IQUE	0	1	1	1
ISTE	0	1	1	1

C'est alors à partir du schéma dérivationnel initial et de la matrice de corrélation locale M_D^{IC} qu'il est procédé à la génération des sous-schémas dérivationnel définitifs.

Nous obtenons dans le cas présent :

$$D_1 = < ERE >$$
$$D_2 = < IE,\ IQUE,\ ISTE >$$

d'où les familles de mots :

$$F_1 = < CHIMERE >$$
$$F_2 = < CHIMIE,\ CHIMIQUE,\ CHIMISTE >$$

Nous trouvons figure 79 un extrait des familles primitives scindées de cette manière.

IV 3. 3. Obtention des familles de mots suffixales complètes :

La "non exhaustivité" de l'ensemble **S** des suffixes (surcomposition des suffixes) nous amène, à ce stade du traitement à être en présence de familles incomplètes :

Exemples :

$$\begin{cases} F_1(ACH) = < ACH + AT,\ ACH + ETER > \\ F_2(ACHET) = < ACHET + ER,\ ACHET + EUR > \end{cases}$$

$$\begin{cases} F_3(AMID) = < AMID + ON,\ AMID + ONNER > \\ F_4(AMIDONN) = < AMIDONN + ER,\ AMIDONN + ERIE > \end{cases}$$

$$\begin{cases} F_5(ART) = < ART,\ ART + ISTE > \\ F_6(ARTIST) = < ARTIST + E,\ ARTIST + IQUE > \end{cases}$$

L'obtention des familles complètes peut être réalisée, ainsi que nous pouvons le voir, par une simple réunion transitive. Cela ne peut cependant être entrepris sans quelques précautions.

150	ALI#EN	* ATION	ER	#E	ISTE		
151	ALIGN	* EMENT	ER				
152	ALIMENT	*	AIRE	ATION	ER		
153	ALL	* ER	URE				
154	ALLAIT	* EMENT	ER				
155	ALL#EG	* ATION					
156	ALL#EG	* EMENT	ER				
157	ALLI	* AGE	ER				
158	ALLI	* ANCE	ER				
159	ALLONG	* E	ER				
160	ALLUM	* AGE	ER	ETTE	EUR	EUSE	OIR
161	ALPIN	*					
162	ALPIN	* ISME	ISTE				
163	ALT	* ESSE					
164	ALT	* ITUDE					
165	ALTERN	* ATEUR	ER				
166	ALTERNAT	* EUR	IF	IVE			
167	ALT#ER	* ATION	ER				
168	ALTRU	* ISME	ISTE				
169	AMADOU	*					
170	AMADOU	* ER					
171	AMAIGR	* IR	ISSEMENT				
172	AMANT	*	E				
173	AMARR	* AGE	ER				
174	AMAZON	* E	ITE				
175	AMBASSAD	* E	EUR				
176	AMBI	* ANCE	ANT				
177	AMBI	* TION					
178	AMBITI	* EUX	ON				
179	AMBR	* E	#E				
308	ARROS	* AGE	ER	OIR			
309	ART	*	ISTE				
310	ART	* ƏERE					
311	ARTICUL	* ATION	ER				
312	ARTIFIC	* E	IER				
313	ARTIFICI	* EL					
314	ARTILL	* ERIE	EUR				
315	ARTISAN	*	AL	AT			
316	ARTIST	* E	IQUE				

fig. 79. Extrait des familles primitives scindées.

IV. 3. 3. 1. Notion de polysémie :

a) Définition : *"polysémie : caractère d'un signe qui possède plusieurs contenus, plusieurs valeurs".*

Petit Robert

b) Propriété :

Soit $\mathbf{F} = \langle F_1, F_2, \ldots, F_n \rangle$ l'ensemble des familles de mots.

Dans ce cadre un polysème m appartient à au moins deux familles différentes notons les F_i et F_j :

$$m \in F_i \quad \text{et} \quad m \in F_j$$

Désignons par r_i (respectivement r_j) le radical associé à la famille F_i (respectivement F_j).

Nous introduirons, à ce stade du traitement, la distinction suivante :

a) si $r_i = r_j$
alors m sera dit polysème ;

b) si $r_i \neq r_j$
alors m sera dit pseudo-polysème.

IV. 3. 3. 2. Réunion transitive :

L'obtention des familles suffixales complètes se fera par réunion transitive autour des pseudo-polysèmes. Ainsi,

$F_1(r_1) = \langle m^1_1 m^1_2 \ldots m^1_p \rangle$ et $F_2(r2) = \langle m^2_1 m^2_2 \ldots m^2_q \rangle$

seront réunies si on peut trouver i et j tels que $m^1_i = m^2_j$ et si $r_1 \neq r_2$.

On aura donc :

$$F_1(r_1) \cup F_2(r_2) = \langle m^1_1 m^1_2 \ldots m^1_p m^2_1 m^2_2 \ldots m^2_q \rangle$$

$$= \langle r_1 + s^1_1, \ldots, r_1 + s^1_p + r_2 + s^2_1, \ldots,$$

$$r_2 + s^2_q \rangle$$

Le schéma dérivationnel associé à $F_1(r_1) \cup F_2(r_2)$ sera :

$$D(r_1) \cup D(r_2) = \langle s^1_1, \ldots, s^1_p, s^2_1, \ldots, s^2_q \rangle$$

Afin d'éviter certaines réunions illicites, on introduit la contrainte supplémentaire suivante : $D(r_1) \cup D(r_2)$ ne doit pas comporter d'incompatibilités.

Exemples :

$$F_1(CHRONOLOG) = \langle CHRONOLOG + IE,\ CHRONOLOG + IQUE \rangle$$
$$F_2(CHRONOLOGIQU) = \langle CHRONOLOGIQU + E,\ CHRONOLOGIQU + EMENT \rangle$$

par réunion transitive on obtient :

$$F_1(CHRONOLOG) \cup F_2(CHRONOLOGIQU) = \langle CHRONOLOG + IE,\ CHRONOLOG + IQUE,\ CHRONOLOG + IQUEMENT \rangle$$

Notons le cas particulier des mots qui sont simultanément polysèmes et pseudo-polysèmes.

Exemples:

$$F_1(SOMM) = \langle SOMM + ATION,\ SOMM + E,\ SOMM + ER \rangle$$
$$F_2(SOMME) = \langle SOMME \rangle$$
$$F_3(SOMME) = \langle SOMME,\ SOMME + IL,\ SOMME + ILLER \rangle$$

La réunion transitive autour de ces mots doit se faire en choisissant manellement, parmi les familles correspondant à un même radical, la famille sur laquelle doit précisément se faire la réunion transitive. On pourrait essayer d'automatiser dans certains cas ce choix par une méthode faisiant de nouveau intervenir les matrices de compatibilités des suffixes.

IV. 3. 3. 3. Choix du radical :

Soit une famille obtenue après réunion transitive des familles $F_1(r_1)$ $F_2(r_2)$, ..., $F_n(r_n)$:

$$F = F_1(r_1) \cup F_2(r_2) \ldots F_n(r_n)$$

On peut écrire :

$$F = \langle m_1^1 \ldots m_{p1}^1 \quad m_1^2 \ldots m_{p2}^2 \quad \ldots m_1^n \ldots m_{pn}^n \rangle$$
$$= \langle r_1 + s_1^1, \ldots, r_1 + s_{p1}^1, r_2 + s_1^2, \ldots, r_2 + s_{p2}^2,$$
$$r_n + s_1^n, \ldots, r_n + s_{pn}^n \rangle$$

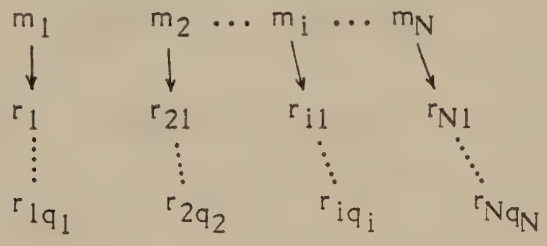
Ainsi à chaque mot $m^k_{ik} \in F$ il correspond un radical r_k. Plusieurs radicaux sont cependant associés aux pseudo-polysèmes.

On peut écrire, après élimination des mots redondants dans F :

$F = < m_1\ m_2\ \ldots\ m_N >$ avec $N <$ card F_i

A chaque m_i il correspond un radical ou plusieurs si m_i est pseudo-polysème.

On a alors :



La solution la plus naturelle pour obtenir le radical r à associer à F est de choisir le radical le plus court. Toutefois, la décomposition multiple de certains mots (pseudo-polysèmes) engendre parfois des radicaux faux trop courts.

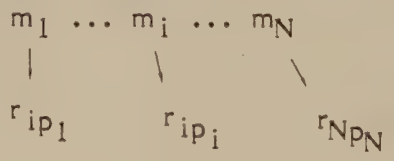
Exemples :

$F_1(RAB) = < RAB + OT,\ RAB + OTER >$

$F_2(RABOT) = < RABOT,\ RABOT + AGE,\ RABOT + ER >$

Ainsi, pour tenir compte de ce phénomène, lorsque tous les mots correspondant au radical le plus court sont des pseudo-polysèmes, ce radical n'est pas retenu.

Le choix du radical (cf. fig. 80 et 82) se fait en deux étapes : 1) dans la première, on associe à chaque mot un seul radical. Celui-ci correspond au radical le plus long. On a alors :



2) le radical r que l'on associera à la famille F est par suite obtenu en prenant le radical r_{ip_i} le plus court.

Exemples :

Reprenons les familles :

$F_1(RAB) = < RAB + OT,\ RAB + OTER >$

$F_2(RABOT) = < RABOT,\ RABOT + AGE,\ RABOT + ER,\ RABOT + EUSE >$

On a alors :

RABOT RABOTAGE RABOTER RABOTEUSE
RAB RABOT RAB RABOT
RABOT RABOT

La première étape consiste à associer à chaque mot le radical le plus long (radicaux soulignés), la deuxième à prendre pour radical le radical le plus court parmi ceux-ci, soit RABOT.

IV. 3. 3. 4. Notion de suffixes composés :

Soit F une famille de mots obtenue après réunion transtive et r le radical associé :

$$F(r) = < m_1, m_2, \ldots, m_N >$$

On peut alors écrire :

$$F(r) = < r + s_1, r + s_2, \ldots, r + s_N >$$

Une chaîne suffixale s_i constituera un suffixe composé si elle n'appartient pas à l'ensemble des suffixes **S** .

Exemple :

F(CAUTI) = < CAUTI + ON, CAUTI + ONNEMENT, CAUTI + ONNER >
ONNEMENT ∉ S est un suffixe composé.

Nous donnons figure 81 la liste des suffixes et suffixes composés ainsi obtenus accompagnés de leur nombre d'occurrences.

La notion de suffixe composé est une justification a posteriori de l'option que nous avons prise, d'accepter les découpages multiples d'un mot qui sont une conséquence de l'imbrication des suffixes.

L'obtention des mêmes familles aurait demandé, en effet, que l'on répertorie comme suffixes les suffixes composés, tâche que nous ne pouvons envisager de faire en raison du travail manuel que cela demanderait.

Nous pouvons remarquer que les suffixes composés obtenus sont des compositions de suffixes :

EMENTAIRE correspond à EMENT + AIRE
ARDER " " ARD + ER
IQUEMENT " " IQUE + MENT
IENNAGE " " IEN + (N) + AGE

Le dernier exemple montre en outre que certaines altérations morphologiques telles que le dédoublement de certaines consonnes sont prises en compte.

Nous trouvons figure 80 un extrait des familles de mots obtenues après réunion transitive. On peut en particulier remarquer, en comparant cette figure à la figure 79, que les familles 309 et 316 de la figure 79 ont été ainsi automatiquement réunies pour donner la famille 279 de la figure 80.

Cette réunion a été accompagnée, comme on peut le voir, par le choix du radical "ART", et donc également par la création du suffixe composé "ISTIQUE" qui n'avait pas été initialement répertorié.

IV. 3. 3. 5. Choix du représentant :

Les familles de mots pourraient être repérées par leur numéro d'ordre (86).

Nous avons choisi, pour des raisons de clarté et de commoditié, de repérer une famille de mots donnée par un mot pris précisément dans la famille elle-même.

Ainsi la famille :

< *ACHAT, ACHETER, ACHETEUR, RACHAT, RACHETER* >

pourrait être représentée par l'un quelconque de ses éléments.

Deux familles ne doivent cependant pas avoir un représentant identique. Aussi, l'élément choisi comme représentant d'une famille donnée ne doit-il pas être polysémique.

Pour permettre au niveau des applications une représentation canonique plus parlante et une gestion plus facile des familles, on tente de choisir pour représentant le terme "originel" ou le plus répandu (ACHETER ou ACHAT dans l'exemple pris ci-dessus).

Deux critères s'offrent à nous pour opérer ce choix. Le premier critère de choix consiste à retenir l'élément le plus court de la famille. Cela se fonde sur le fait que la plupart des familles se sont formées par dérivation (en général non régressive) et/ou par composition, ce qui allonge ordinairement le mot primitif. Ce critère n'est pas non plus sans rappeler l'hypothèse selon laquelle les mots tendent à se "raccourcir" lorsque leur fréquence d'utilisation augmente. Le deuxième critère consiste à choisir comme représentant d'une famille, le mot dont le suffixe a la plus grande fréquence dans le lexique. Une combinaison des deux critères a été utilisée pour le choix des représentants (cf. fig. 82). Le deuxième critère devrait cependant donner les mêmes résultats si le calcul des fréquences suffixales était effectué sur un corpus.

32	ACCESSOIR	* E	EMENT			
33	ACC*EL*ER	* ATION	ER			
34	ACCIDENT	*	EL	ER		
35	ACCLAM	* ATION	ER			
36	ACCLIMAT	* ATION	ER			
37	ACCOMPAGN	* EMENT	ER			
38	ACCOMPL	* IR	ISSEMENT			
39	ACCORD	*	ER	EUR		
40	ACCOST	* AGE	ER			
41	ACCOUCH	* EMENT	ER			
42	ACCOUTR	* EMENT	ER			
43	ACCROCH	* AGE	ER			
44	ACCUMUL	* ATEUR	ATRICE	ER		
45	ACCUS	* ATEUR	ATION	ATRICE	ER	#E
46	ACCUSAT	* IF				
47	ACH	* AT	ETER	ETEUR		
48	ACHALAND	* AGE	#E			
49	ACHARN	* EMENT	ER			
50	ACHEMIN	* EMENT	ER			
51	ACHOPP	* EMENT	ER			
52	ACID	* E	#E	IT#E		
53	ACQU*ER	* EUR	IR			
54	ACQUIESC	* EMENT	ER			
55	ACQUITT	* EMENT	ER			
56	ACROBAT	* E	IE			
57	ACT	* E	IF			
58	ACT	* EUR				
59	ACTI	* ON	ONNER			
60	ACTIONN	* AIRE				
269	AROM	* ATE	ATIQUE	E		
270	ARPENT	*	AGE	ER	EUR	
271	ARR-	* ET	ET#E			
272	ARRACH	* AGE	ER			
273	ARRANG	* EMENT	ER			
274	ARRIM	* AGE	ER			
275	ARRIV	* AGE	ER	#EE		
276	ARROG	* ANCE	ER			
277	ARROND	* IR	ISSEMENT			
278	ARROS	* AGE	ER	OIR		
279	ART	*	ISTE	ISTIQUE		
280	ART	* @ERE				

fig. 80. Extrait des familles de mots obtenues après réunion transitive.

1	¬ATRE	11	61	ANTEUR	1	121	ELEUR	1	181	ICULE	7	241	ITUDE	21	301	SER	2
2	¬ATRER	1	62	ANTIN	1	122	ELME	2	182	IE	107	242	IVE	9	302	SION	17
3	¬ATRETRE	1	63	ANTINE	1	123	ELIER	1	183	IEL	1	243	IVEMENT	1	303	TAIRE	2
4	¬ATRIE	1	64	ARD	41	124	ELIDERE	1	184	IENENT	1	244	IVETRE	1	304	TATION	1
5	ABILITRE	9	65	ARDE	2	125	ELLE	44	185	IEN	45	245	IVITRE	1	305	TER	2
6	ABLE	107	66	ARDER	2	126	ELLEMENT	2	186	IENNAGE	1	246	LADE	1	306	TRE	24
7	ABLEMENT	2	67	ARDIERE	1	127	ELURE	3	187	IENNE	35	247	LER	1	307	TION	51
8	ACE	9	68	ARDISE	1	128	EMENT	340	188	IER	235	248	LISER	1	308	TIONNAIRE	1
9	ACER	1	69	AS	17	129	EMENTAIRE	2	189	IEUX	1	249	LITRE	1	309	TIONNEMENT	1
10	ACRE	3	70	ASSE	8	130	EMENTAIREMENT	1	190	IERE	116	250	MENT	57	310	TIONNER	1
11	ADE	51	71	ASSER	21	131	EMENTAL	1	191	IRE	4	251	MENTAIRE	1	311	TOIRE	14
12	ADER	1	72	ASSERIE	1	132	EMENTATION	1	192	IF	54	252	MENTATION	1	312	TOIREMENT	2
13	ADIE	1	73	ASSNE	1	133	EMENTER	2	193	IFIABLE	1	253	MENTER	1	313	U	45
14	ADIER	1	74	ASSIER	1	134	EMENTERIE	1	194	IFIER	36	254	NICHER	1	314	UAIRE	1
15	ADIF	1	75	ASSON	1	135	ENCE	41	195	IFLER	1	255	OCHAGE	1	315	UE	7
16	ADIN	1	76	AT	56	136	EOIRE	1	196	IL	18	256	OCHER	7	316	UEUSE	1
17	AGE	242	77	ATAIRE	6	137	EON	1	197	ILLAGE	3	257	OIR	60	317	UEUX	10
18	AGEABLE	1	78	ATATION	1	138	EOYER	1	198	ILLANT	1	258	OIRE	18	318	ULE	14
19	AGEOIS	1	79	ATE	1	139	ER	1693	199	ILLE	34	259	OIS	7	319	URAL	1
20	AGER	6	80	ATEMENT	1	140	ERAIE	2	200	ILLER	35	260	OISE	1	320	URATION	1
21	AGEUX	1	81	ATER	1	141	EREAU	4	201	ILLEUX	3	261	OISERIE	1	321	URE	111
22	AGERE	1	82	ATEUR	91	142	ERELLE	3	202	ILLNE	1	262	OLE	12	322	URER	1
23	AGERement	1	83	ATIF	11	143	ERESQUE	1	203	ILLON	18	263	ON	240	323	URERIE	3
24	AGINE	1	84	ATION	367	144	ERESSE	7	204	ILLONNER	2	264	ONISTE	1	324	URIER	3
25	AGISTE	1	85	ATIQUE	14	145	ERETTE	1	205	IN	101	265	ONNABLE	2	325	URIERE	1
26	AIE	4	86	ATISER	1	146	ERIE	148	206	INAIRE	1	266	ONNABLEMENT	1	326	URITRE	1
27	AIL	18	87	ATIVE	2	147	ERISER	1	207	INAISON	1	267	ONNADE	1	327	UTION	1
28	AILLANT	1	88	ATOIRE	11	148	EROLE	1	208	INAL	1	268	ONNAGE	3	328		803
29	AILLE	17	89	ATOIREMENT	1	149	ERON	11	209	INE	89	269	ONNAIRE	6			
30	AILLER	18	90	ATRICE	31	150	ESCUE	4	210	INEMENT	1	270	ONNE	4			
31	AILLEUX	1	91	ATURE	15	151	ESSE	43	211	INER	34	271	ONNEL	4			
32	AIN	32	92	ATURER	1	152	ET	115	212	INET	1	272	ONNEMENT	9			
33	AINE	22	93	AUD	9	153	ETAGE	1	213	INETTE	1	273	ONNER	102			
34	AINEMENT	1	94	AUDE	2	154	ETER	26	214	IOLE	4	274	ONNERIE	3			
35	AIRE	112	95	AUT	3	155	ETERIE	2	215	ION	4	275	ONNET	1			
36	AIREMENT	6	96	BILISATEUR	1	156	ETEUR	1	216	IQUE	139	276	ONNETTE	2			
37	AIS	11	97	BILISATION	1	157	ETRE	18	217	IQUEMENT	9	277	ONNEUR	2			
38	AISON	15	98	BILISER	1	158	ETIER	2	218	IR	168	278	ONNEUSE	1			
39	AL	82	99	BILITRE	31	159	ETIDERE	1	219	IS	54	279	ONNEUX	2			
40	ALE	6	100	BLE	47	160	ETIS	2	220	ISABLE	2	280	ONNIER	5			
41	ALEMENT	5	101	BLEMENT	1	161	ETTE	135	221	ISATEUR	2	281	ONNIDERE	2			
42	ALISATION	1	102	BUS	1	162	ETTERIE	1	222	ISATION	16	282	ONNISTE	1			
43	ALISER	2	103	CEUR	1	163	EUL	1	223	ISE	36	283	OSE	11			
44	ALISME	3	104	CIR	4	164	EUR	302	224	ISER	84	284	OT	46			
45	ALISTE	1	105	CISSEMENT	1	165	EURE	3	225	ISME	58	285	OTE	15			
46	ALITRE	8	106	CULE	1	166	EURRE	1	226	ISON	4	286	OTER	27			
47	AN	22	107	E	1329	167	EUSE	110	227	ISSABLE	1	287	OTIQUE	1			
48	ANCE	84	108	EABLE	1	168	EUX	150	228	ISSAGE	6	288	OTISME	1			
49	ANCEMENT	1	109	EADE	1	169	DERE	38	229	ISSANT	1	289	OTTE	10			
50	ANCER	1	110	EAGE	1	170	ME	266	230	ISSEMENT	56	290	OTTER	7			
51	ANCIER	3	111	EANT	2	171	MEAT	1	231	ISSERIE	1	291	OTTERIE	1			
52	ANDE	7	112	EAT	1	172	MEATE	1	232	ISSEUR	6	292	OUILLER	4			
53	ANDER	1	113	EAU	89	173	MEE	92	233	ISSEUSE	2	293	OYANT	1			
54	ANDIER	3	114	EAUD	1	174	MEEN	2	234	ISTE	88	294	OYER	19			
55	ANDIERE	2	115	EINE	1	175	MEMENT	3	235	ISTERIE	2	295	RE	52			
56	ANE	12	116	EL	47	176	M	5	236	ISTIQUE	4	296	REMENT	1			
57	ANT	231	117	ELABLE	1	177	NABLE	1	237	ITAIRE	2	297	SABLE	1			
58	ANTE	23	118	ELER	26	178	NAL	3	238	ITE	24	298	SATEUR	1			
59	ANTER	1	119	ELET	9	179	NBILITRE	1	239	ITRE	119	299	SATION	2			
60	ANTERIE	1	120	ELETTE	2	180	NBLE	18	240	ITION	35	300	SATRICE	1			

fig. 81. Ensemble des suffixes et suffixes composés.

Radical	Représentant		Eléments des familles de mots			
ACCEPT	ACCEPTER.	*	ACCEPTATION	ACCEPTER	ACCEPTEUR	
ACCESSI	ACCESSION.	*	ACCESSIBLE	ACCESSION		
ACCESSOIR	ACCESSOIRE.	*	ACCESSOIRE	ACCESSOIREMENT		
ACC#EL#ER	ACC#EL#ERER.	*	ACC#EL#ERATION	ACC#EL#ERER		
ACCIDENT	ACCIDENT.	*	ACCIDENT	ACCIDENTEL	ACCIDENTER	
ACCLAM	ACCLAMER.	*	ACCLAMATION	ACCLAMER		
ACCLIMAT	ACCLIMATER.	*	ACCLIMATATION	ACCLIMATER		
ACCOMPAGN	ACCOMPAGNER.	*	ACCOMPAGNEMENT	ACCOMPAGNER		
ACCOMPL	ACCOMPLIR.	*	ACCOMPLIR	ACCOMPLISSEMENT		
ACCORD	ACCORD.	*	ACCORD	ACCORDER	ACCORDEUR	
ACCOST	ACCOSTER.	*	ACCOSTAGE	ACCOSTER		
ACCOUCH	ACCOUCHER.	*	ACCOUCHEMENT	ACCOUCHER		
ACCOUTR	ACCOUTRER.	*	ACCOUTREMENT	ACCOUTRER		
ACCROCH	ACCROCHER.	*	ACCROCHAGE	ACCROCHER		
ACCUMUL	ACCUMULER.	*	ACCUMULATEUR	ACCUMULATRICE	ACCUMULER	
ACCUS	ACCUSER.	*	ACCUSATEUR	ACCUSATION	ACCUSATRICE	ACCUSER
		*	ACCUS#E			
ACCUSAT	ACCUSATIF.	*	ACCUSATIF			
ACH	ACHAT.	*	ACHAT	ACHETER	ACHETEUR	
ACHALAND	ACHALANDAGE.	*	ACHALANDAGE	ACHALAND#E		
ACHARN	ACHARNER.	*	ACHARNEMENT	ACHARNER		
ACHEMIN	ACHEMINER.	*	ACHEMINEMENT	ACHEMINER		
ACHOPP	ACHOPPER.	*	ACHOPPEMENT	ACHOPPER		
ACID	ACIDE.	*	ACIDE	ACID#E	ACIDIT#E	
ACQU#ER	ACQU#ERIR.	*	ACQU#EREUR	ACQU#ERIR		
ACQUIESC	ACQUIESCER.	*	ACQUIESCEMENT	ACQUIESCER		
ACQUITT	ACQUITTER.	*	ACQUITTEMENT	ACQUITTER		
ACROBAT	ACROBATE.	*	ACROBATE	ACROBATIE		
ACT	ACTE.	*	ACTE	ACTIF		
ACT	ACTEUR.	*	ACTEUR			
ACTI	ACTION.	*	ACTION	ACTIONNER		

fig. 82. Exemples de familles suffixales obtenues automatiquement.
(Extrait du dictionnaire)

Remarque : Nous faisons suivre d'un point le mot qui sert à repérer une famille de mots donnée. Nous marquons ainsi qu'il s'agit du représentant d'un ensemble (cf. § X).

V. Traitement des préfixes :

On peut effectuer les mêmes traitements sur les préfixes.

Les préfixes peuvent cependant se rencontrer aussi bien dans des verbes, que dans des substantifs ou adjectifs.

Aussi, une famille primitive sera telle que :

$$F = \langle m_i \mid m_i = p_i + r \qquad m_i \in \mathbf{M},\ p_i \in \mathbf{P} \rangle$$

Elle sera dite préfixale, car obtenue par décomposition en préfixe + "radical", par opposition aux familles précédentes qui sont suffixales.

Nous donnons figure 83 un extrait de quelques familles ainsi obtenues.

Leur nombre étant relativement peu important (1 800 familles environ), leur correction a été faite manuellement.

VI. Obtention des familles de mots complètes :

L'obtention des familles complètes, répondant à la définition que nous avons donnée au paragraphe IV. 1., s'effectue en opérant à nouveau une réunion transitive entre les familles suffixales et les familles préfixales.

Exemples :

Soit les familles préfixales :

$$F_1(ACHAT) = \langle ACHAT,\ RACHAT \rangle$$
$$F_2(ACHETER) = \langle ACHETER,\ RACHETER \rangle$$

et la famille suffixale :

$$F_3(ACH) = \langle ACHAT,\ ACHETER,\ ACHETEUR \rangle$$

Par réunion transistive on obtient :

$$F(ACH) = \langle ACHAT,\ ACHETER,\ ACHETEUR,\ RACHAT,\ RACHETER \rangle$$

Là encore, un traitement "manuel" des polysèmes s'impose.

30	ADMIRER	*	ADMIRER	ENTRADMIRER		
31	ADMISSIBLE	*	ADMISSIBLE	INADMISSIBLE		
32	ADOSSEMENT	*	ADOSSEMENT	SURADOSSEMENT		
33	ADRESSE	*	ADRESSE	MALADRESSE		
34	ADROIT	*	ADROIT	MALADROIT		
35	AFFECTATION	*	AFFECTATION	D#ESAFFECTATION	R#EAFFECTATION	
36	AFFECTER	*	AFFECTER	D#ESAFFECTER		
37	AFFERMIR	*	AFFERMIR	RAFFERMIR		
38	AFFINAGE	*	AFFINAGE	RAFFINAGE		
39	AFFINER	*	AFFINER	RAFFINER		
40	AFFOLER	*	AFFOLER	RAFFOLER		
41	AGE	*	AGE PARAGE	ADAGE RAGE	FORAGE TRIAGE	IMAGE
42	AGER	*	IMAGER	RAGER	VIAGER	
43	AGIR	*	AGIR	R#EAGIR		
44	AGR#EABLE	*	AGR#EABLE	D#ESAGR#EABLE		
45	AGR#EMENT	*	AGR#EMENT	D#ESAGR#EMENT		
46	AIDE	*	AIDE	ENTRAIDE	RAIDE	
47	AIDER	*	AIDER	ENTRAIDER		
48	AIL	*	AIL	CORAIL	RAIL	
49	AILLER	*	CISAILLER	RAILLER		
50	AIN	*	ENTRAIN	FORAIN		
51	AISE	*	AISE	ALAISE	MALAISE	
52	AISER	*	APAISER	BIAISER		
53	AIS#E	*	AIS#E	MALAIS#E		
54	AJOUTER	*	AJOUTER	RAJOUTER		
55	AJUSTEMENT	*	AJUSTEMENT	RAJUSTEMENT		
56	AJUSTER	*	AJUSTER	RAJUSTER		
57	AL	*	P#ENAL	POSTAL		
58	ALIMENTATION	*	ALIMENTATION	R#EALIMENTATION		

fig. 83. Extrait de quelques familles primitives préfixales.

VII. Gestion des familles de mots :

La constitution d'un tel dictionnaire de familles de mots nous amène au problème traditionnel inhérent à la maintenance des dictionnaires et des fichiers en général.

Pour cela, nous avons été amenés à définir un langage de commande permettant les opérations suivantes :

- Edition : cette opération devant permettre l'édition d'une ou plusieurs familles.
- Mise à jour : cette opération comprend plusieurs commandes dont les principales sont :
  a) création d'une nouvelle famille,
  b) suppression d'une famille,
  c) opérations internes aux familles de mots. Celles-ci peuvent être des opérations d'ajout, suppression ou remplacement d'éléments à l'intérieur d'une même famille,
  d) opérations de changement d'un représentant ou d'un radical.

Notons la nécessité d'un dictionnaire inverse (mot - représentant) si l'on veut rendre aisées les opérations décrites ci-dessus. En effet, un tel fichier permet, notamment, d'accéder à une famille par l'un quelconque de ses éléments. Cette possibilité sera d'ailleurs d'autant plus nécessaire si l'on songe à l'extention de ce dictionnaire en y incluant, par exemple, des ensembles de mots dont la parenté sémantique n'est pas basée sur une proximité morphologique (**ex.** synonymie : LOGICIEL - SOFTWARE).

Nous donnons figure 84 quelques exemples de manipulation des familles de mots.

Du fait de la polysémie, remarquons, qu'outre l'analyse "syntaxique" de l'instruction associée à l'opération que l'on veut effectuer, qu'une analyse "sémantique", en quelque sorte, de cette instruction est nécessaire avant de lancer l'exécution effective de celle-ci (cf. la quatrième instruction de la figure 84).

VIII. Bilan de l'expérience :

Les traitements ont été effectués sur un dictionnaire initial de 23 494 mots. Celui-ci correspond en fait à 170 000 formes dérivées différentes (conjugaison des verbes, féminin des adjectifs, etc.) qui peuvent évidemment être associées à leurs formes primitives.

```
HEURE : 08 H 57 MN          DATE : 07 / 06 / 79          MONITEUR DE GESTION DES FAMILLES DE MOTS
*****************************************************************************************************
*****************************************************************************************************
*****************************************************************************************************
I N S T R U C T I O N : EDIT:AMBULANT.

REPRESENTANT : AMBULANT.                    RADICAL : AMBULANT
                    AMBULANCE
                    AMBULANT
                    AMBULANCIER
                    AMBULANTE
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : AMBULANT.:=AMBULANT.-AMBULANCE-AMBULANCIER

REPRESENTANT : AMBULANT.                    RADICAL : AMBULANT
                    AMBULANT
                    AMBULANTE
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : AMBULANCE.:=AMBULANC+AMBULANCE+AMBULANCIER

REPRESENTANT : AMBULANCE.                   RADICAL : AMBULANC
                    AMBULANCE
                    AMBULANCIER
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : BOUCHER*:=BOUCHER*+D#EBOUCHER*/

***CREATION (OU MODIFICATION) N' EST PAS EFFECTUEE (MOT POLYSEMIQUE)***

REPRESENTANT : BOUCHERIE.                   RADICAL : BOUCH
                    BOUCHER
                    BOUCHERIE
-----------------------------------------------------------------------------------------------------
REPRESENTANT : BOUCHON.                     RADICAL : BOUCH
                    BOUCHER
                    BOUCHON
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : EDIT:D#EBOUCHER*

REPRESENTANT : D#EBOUCHER.                  RADICAL : D#EBOUCH
                    D#EBOUCHAGE
                    D#EBOUCHER
                    D#EBOUCH#E
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : BOUCHON.:=BOUCHON.+D#EBOUCHER*-D#EBOUCH#E

REPRESENTANT : BOUCHON.                     RADICAL : BOUCH
                    BOUCHER
                    BOUCHON
                    D#EBOUCHAGE
                    D#EBOUCHER
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : D#EBOUCHER.:=D#EBOUCHER.-D#EBOUCHAGE

REPRESENTANT : D#EBOUCHER.                  RADICAL : D#EBOUCH
                    D#EBOUCH#E
                    D#EBOUCHER
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : LOGICIEL.:=SSS+LOGICIEL+SOFTWARE

REPRESENTANT : LOGICIEL.                    RADICAL : SSS
                    LOGICIEL
                    SOFTWARE
-----------------------------------------------------------------------------------------------------
I N S T R U C T I O N : LOGICIEL.=SOFTWARE.

I N S T R U C T I O N : SOFTWARE.:=XXX+SOFTWARE.

REPRESENTANT : SOFTWARE.                    RADICAL : XXX
                    LOGICIEL
                    SOFTWARE
-----------------------------------------------------------------------------------------------------
```

fig. 84. Quelques exemples de gestion des familles de mots.

L'élimination de ce dictionnaire des mots qui ne peuvent a priori participer à des familles de mots (prépositions, articles, conjonction, pronom, ..., noms propres, ...) nous permet d'obtenir un dictionnaire de 19 432 mots qui, précisément, est soumis aux différents traitements d'obtention des familles de mots.

Les familles primitives suffixales obtenues, au nombre de 4 434, ont fait apparaître un taux d'erreur (familles incorrectes) de l'ordre de 40 %.

Il est intéressant de noter que seulement 11 023 mots ont participé à la définition des familles primitives suffixales. Autrement dit, à ce stade, 8 000 mots environ constituent des singletons, c'est à dire qu'ils forment à eux seuls leur propre famille. Ce nombre devrait cependant baisser sensiblement si l'on remarque que dans ces comptages nous n'avons pas retenu les mots à préfixe (1 500 environ participent à des familles préfixales correctes) d'une part, et les mots à suffixes composés d'autre part (par exemple la famille < PREVISION, PREVISIONNEL, PREVISIONNELLEMENT > ne pourra être obtenue que si l'on tient compte des suffixes composés).

L'explosion des familles suffixales primitives au moyen des matrices de compatibilité des suffixes engendre 5 440 familles. Celles-ci ne présentent plus que 5 % environ de familles suffixales incorrectes. 250 familles suffixales ont ainsi subi une correction manuelle.

Nous donnons figure 85(a) l'évolution du nombre de familles suffixales au cours des diverses phases du traitement. La figure 85(c) donne l'organigramme général de l'algorithme qui a été utilisé.

Traitement	Familles primitives	Explosion	Correction	Réunion transitive
Nbre de fam. suffixales	4 434	5 440	5 938	5 280

fig. 85(a).

Enfin la répartition des règles de compatibilité et d'incompatibilité des suffixes se présente comme suit :

	Mat. de corrélation des suffixes M	Mat. des compatibilités M_C	Mat. des incompatibilités M_I	Points ambigus
Nbre de pts des matrices	2 639	404	1 548	687
Nbre de couples de mots correspondant	13 795	1 669	2 486	9 640

fig. 85(b).

Les chiffres donnés figure 85(b) correspondent aux demi-matrices puisque celles-ci sont symétriques.

Les remarques suivantes peuvent être faites :

a) le nombre des règles d'incompatibilité est le plus important (58,66 %) mais il ne correspond qu'à 18 % des couples de mots.

b) les points que nous avons appelés ambigus dans le texte et dont l'incompatibilité est contextuelle (fait intervenir le schéma dérivationnel) sont relativement peu nombreux (26 %) mais correspondent en fait à 70 % des couples de mots, autrement dit à la plus grande partie des familles de mots. Cette constatation met en relief la performance de l'algorithme utilisé.

c) une performance meilleure aurait pu être atteinte si l'on avait admis une compatibilité et/ou une incompatibilité partielles, et non totales, de certains suffixes (par exemple le point *ATION - ATEUR* est 74 fois compatible et 1 fois incompatible). Un tel choix aurait cependant conduit ultérieurement à un travail de correction hétérogène.

Notons enfin que l'algorithme proposé est d'autant plus performant que les dictionnaires soumis au traitement sont exhaustifs. Les schémas dérivationnels sont en effet dans ce cas "plus complets", l'idéal étant l'obtention de schémas saturés.

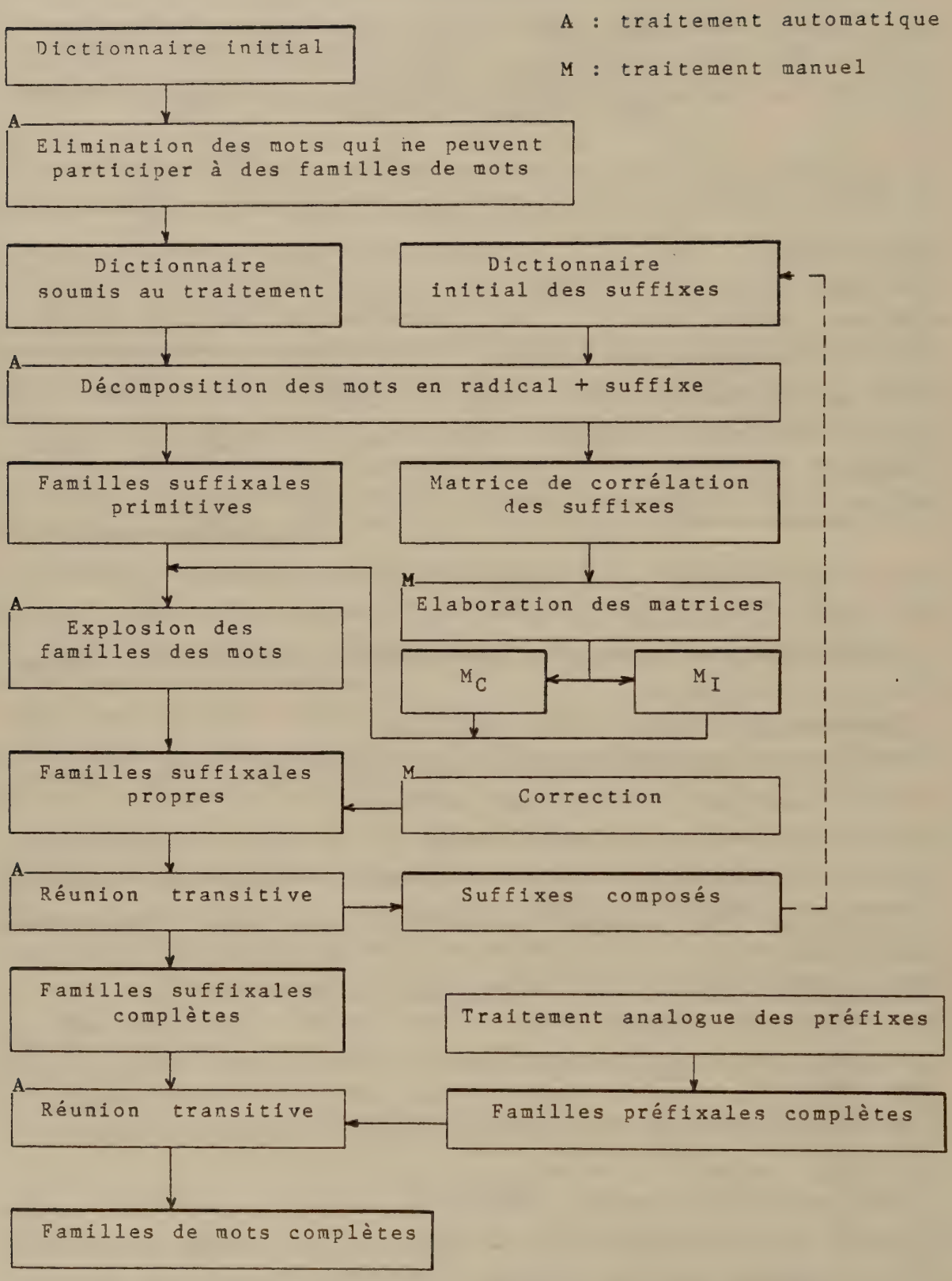


fig. 85.(c) Organigramme général de l'algorithme de construction des familles de mots.

IX. Conclusion :

Nous avons exposé dans ce qui précède une stratégie permettant d'obtenir des familles de mots à partir d'un dictionnaire de mots accompagnés de leurs catégories grammaticales.

Il est à noter que des sous-schémas dérivationnels tels que < ABLE, AGE, ER > dont on pourrait penser a priori qu'ils n'engendrent que des sous-familles correctes ne vérifient pas dans la pratique cette hypothèse. Dans le cas présent on obtiendrait entre autre :

CURABLE, CURAGE, CURER

La méthode exposée permet précisément, en tenant compte des autres suffixes de détecter les radicaux issus de décompositions illicites.

Retenons également que certains polysèmes ont pu être détectés. Ces informations sont importantes tant du point de vue économique (nombre d'appels des procédures de traitement de la polysémie) que du point de vue du traitement effectif de la polysémie (informations apportées par les autres mots de la famille polysémique). Il reste cependant à répertorier les polysèmes entièrement inclus dans les familles de mots.

Les méthodes exposées pourraient, nous semble-t-il, être utilisées pour d'autres langues dans la mesure où ces langues présentent une analogie avec le français quant à leur comportement morphologique (suffixation et préfixation).

X. Description paradigmatique de la phrase :

Les informations ainsi acquises vont désormais nous permettre d'établir la description paradigmatique de la phrase.

Celle-ci consiste en effet à associer à chacun des mots :

- d'une part, le ou les mots dont ils proviennent. Cette description est classique : il s'agit de faire correspondre aux verbes la ou les formes infinitives dont ils dérivent, aux substantifs leurs formes au singulier et aux adjectifs enfin leurs formes au masculin singulier.
- d'autre part, et cela constitue un deuxième niveau de description, la ou les familles de mots auxquelles ils sont susceptibles d'appartenir. Il s'agit en l'occurrence de faire correspondre à chacun des mots du premier niveau descriptif, le ou les représentants des familles de mots auxquelles potentiellement ils appartiennent.

A chaque mot il correspondra donc deux niveaux descriptifs.

Les occurrences suivantes recevraient par exemple la description paradigmatique qui suit :

	Niveau I	Niveau II[1]
interministérielles	*interministériel*	*ministre.*
loués	*louer*	*location/louange.*[2]
nomination	*nomination*	*nommer.*
ouvrent	*ouvrir/ouvrer*	*ouvrir/ouvrer.*

Soulignons cependant dès maintenant que cette description paradigmatique combinée avec la description syntagmatique (cf. ch. III), confèrera à la phrase une représentation à plusieurs niveaux (cf. chV, § II).

Ainsi que le montrent les exemples, c'est donc au stade de la description paradigmatique que sont mises à jour les ambiguïtés lexicales, de même que sont mises à jour les ambiguïtés syntaxiques lors de la description syntagmatique.

Rappelons cependant que seules sont maintenues les ambiguïtés lexicales qui ne peuvent être résolues lors de l'analyse grammaticale (cf. ch. III, Introduction et ch. V, § V).

La description paradigmatique d'une phrase aura donc l'allure suivante :

Exemples :

	Niveau I	Niveau II
le	*le*	*le*
bail	*bail*	*bail.*
expire	*expirer*	*expirer (finir)/expirer (mourir)/expirer (respirer).*
à	*à*	*à*
la	*la*	*la*
fin	*fin*	*fin.*
du	*du*	*du*
mois	*mois*	*mois.*

---

(1) Afin de distinguer les corrélatifs lexicaux des premier et deuxième niveaux descriptifs et marquer ainsi leur différence, nous faisons suivre les seconds par un point (cf. remarque, fig. 82).

(2) "/" indique le "OU exclusif".

De même :

Expiration du bail à la fin du mois.

recevra la description paradigmatique suivante :

Expiration	*Expiration*	*expirer(finir)/expirer (respirer).*
du	*du*	*du*
bail	*bail*	*bail.*
à	*à*	*à*
la	*la*	*la*
fin	*fin*	*fin.*
du	*du*	*du*
mois	*mois*	*mois.*

La figure suivante (fig. 86) offre quelques exemples de description paradigmatique de phrases extraites des textes de la règlementation de l'urbanisme.

N° du mot	*Mots*	*Valeurs grammaticales*	*GN*	***Niveau I***	***Niveau II***
1497	L'	ARTD	1010	L'	L'
1498	#ECHANGE	S	1000	#ECHANGE	#ECHANGER
1499	NE	NEG NE	0000	NE	NE
1500	PEUT	INDC	1010	POUVOIR	POUVOIR
1501	¬ETRE	INF ETAT	1111	¬ETRE	¬ETRE
1502	IMPOS#E	PPAS AT	1000	IMPOSER	IMPOSER
1503	AUX	PREPART	0101	AUX	AUX
1504	LOCATAIRES	S	0101	LOCATAIRE	LOCATION
1505	QUI	REL SUJ	1111	QUI	QUI
1506	ONT	INDC AUX	0101	AVOIR	AVOIR
1507	#ET#E	PPAS ETAT APR AUX	1000	¬ETRE	¬ETRE
1508	SINISTR#E	AT	1000	SINISTR#E	SINISTRE
1509	DANS	PREP	1111	DANS	DANS
1510	LES	ARTD	0101	LES	LES
1511	LIEUX	S	0100	LIEU	LIEU
1512	LOU#ES	PPAS ADJP	0100	LOUER	LOCATION|LOUANGE
1513	.	.	0000	.	.

PHRASE NO. : 17

LES ASCENSEURS NE SONT PAS CONSID#ER#ES COMME UN MOYEN D' #EVACUATION .

Niv.I LES ASCENSEUR NE ¬ETRE PAS CONSID#ERER COMME UN MOYEN D' #EVACUATION .

Niv.II LES ASCENSEUR NE ¬ETRE PAS CONSID#ERER COMME UN MOYEN D' #EVACUER .

PHRASE NO. : 65

CES SECTIONS , DE CARACT@ERE PERMANENT , DOIVENT COMPRENDRE DOUZE MEMBRES AU MINIMUM ET VINGT MEMBRES AU MAXIMUM .

Niv.I CES SECTION , DE CARACT@ERE PERMANENT , DEVOIR COMPRENDRE DOUZE MEMBRE AU MINIMUM ET VINGT MEMBRE AU MAXIMUM .

Niv.II CES SECTION , DE CARACT@ERE PERMANENT , DEVOIR COMPR#EHENSION\|COMPRENDRE(CONTENIR) DOUZE MEMBRE AU MINIMUN ET VINGT MEMBRE AU MINIMUN .

PHRASE NO. : 89

$00000360 D@ES L' AFFICHAGE @A LA MAIRIE D' UN EXTRAIT DE LA D#ECISION PORTANT PERMIS DE CONSTRUIRE ET JUSQU' @A L' EXPIRATION D' UN D#ELAI D' UN AN ET UN MOIS APR@ES CET AFFICHAGE , TOUTE PERSONNE INT#ERESS#EE PEUT CONSULTER LES PI@ECES SUIVANTES DU DOSSIER : ARR¬ET#E ACCORDANT LE PERMIS , PLAN-MASSE , PLAN D' EX#ECUTION , DEVIS DESCRIPTIF .

Niv.I $00000360 D@ES L' AFFICHAGE @A LA MAIRIE D' UN EXTRAIT DE LA D#ECISION PORTANT PERMIS DE CONSTRUIRE ET JUSQU' @A L' EXPIRATION D' UN D#ELAI D' UN AN ET UN MOIS APR@ES CET AFFICHAGE , TOUT PERSONNE INT#ERESSER POUVOIR CONSULTER LES PI@ECE SUIVANT DU DOSSIER : ARR¬ET#E ACCORDER LE PERMIS , PLAN-MASSE , PLAN D' EX#ECUTION , DEVIS DESCRIPTIF .

Niv.II $00000360 D@ES L' AFFICHER @A LA MAIRE D' UN EXTRAIT DE LA D#ECIDER PORTER PERMETTRE DE CONSTRUIRE ET JUSQU' @A L' EXPIRER(ECHEANCE)\|RESPIRER D' UN D#ELAI D' UN AN ET UN MOIS APR@ES CET AFFICHER , TOUT PERSONNE INT#ERESSER POUVOIR CONSULTER LES PI@ECE SUIVRE DU DOSSIER : ARR¬ET#E ACCORDER(DONNER)\|ACCORDER(UNIR)\|ACCORDOIR LE PERMETTRE , PLAN-MASSE , PLAN D' EX#ECUTER(FAIRE)\|EX#ECUTER(TUER) , DEVIS D#ECRIRE .

fig. 86

Chapitre cinquième

COMPOSANTE SEMANTIQUE DU SYSTEME

ACQUISITION ET UTILISATION DES RELATIONS LEXICALES SEMANTIQUES

Dans les deux chapitres précédents, nous avons, du système, décrit respectivement les composantes syntaxique d'une part et morphologique d'autre part. La première de ces composantes devait nous conduire à la description syntagmatique de la phrase, la seconde à sa description paradigmatique.

Nous allons dans le présent chapitre nous attacher à la description de la troisième et dernière partie du système, à savoir sa composante sémantique.

Celle-ci devra nous conduire cette fois à la résolution des ambiguïtés syntaxiques ou lexicales qui auront été relevées lors des deux étapes précédentes et, par suite, à une description définitive de la phrase.

C'est cette description définitive qui constituera la représentation de la phrase.

Nous distinguerons donc deux parties :

- l'une sera consacrée à la résolution des ambiguïtés syntaxiques rencontrées lors de la reconnaissance des diverses relations de dépendance,
- l'autre à la résolution des ambiguïtés lexicales, c'est à dire celles liées à la polysémie.

L'une et l'autre résolution relèvent cependant des mêmes mécanismes (cf. ch.I). En particulier elles sont toutes deux fondées sur la connaissance de compatibilités sémantiques entre unités lexicales que nous avons appelées RLS (cf. ch.II).

Or, si nous sommes déjà en possession des RLS paradigmatiques (cf. ch. IV), nous ne disposons pas encore par contre des relations fondamentales que sont les RLS syntagmatiques.

Deux problèmes distincts se posent alors préalablement :

- l'un est relatif à l'acquisition automatique des informations de ce deuxième axe,
- l'autre, et nous rejoignons ici les mécanismes de résolution, à celui de leur utilisation.

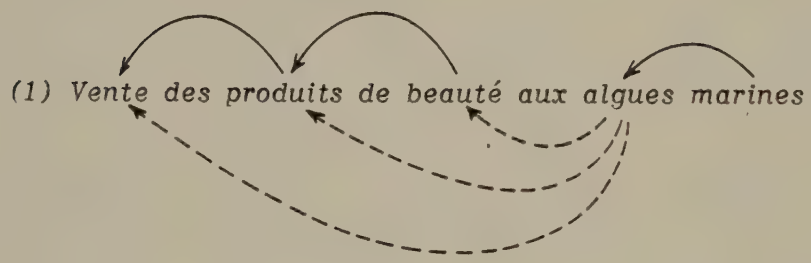
I. Acquisition automatique des RLS syntagmatiques :

I. 1. Hypothèse de base :

Les RLS syntagmatiques que nous avons retenues correspondent, comme nous avons pu le voir, aux relations de dépendance les plus répandues.

Or, si la détermination de celles-ci conduit souvent à des ambiguïtés syntaxiques (nous avons pour cela attribué l'adjectif "potentiel" à toutes les relations déterminées dans le chapitre III), plus souvent encore leur détermination ne se heurte à aucune ambiguïté. Nous entendons par là que plus nombreuses encore sont les relations de dépendance qui peuvent être déterminées en faisant appel seulement à une composante syntaxique. Les nombreux exemples du chapitre III sont à cet égard illustratifs, mais nous n'avions pas, toutefois, mis en relief cet aspect.

Reprenons en effet la phrase :



et considérons parallèlement la phrase suivante :

• Dans la phrase (1) : les informations grammatico-syntaxiques nous amènent à rattacher le substantif "algues" de façon potentielle aux trois substantifs : "beauté", "produits" et "ventes" (flèches en pointillés, cf. fig. 22). Nous sommes donc en présence d'une ambiguïté qui ne peut être résolue que par l'apport d'informations de nature sémantique (compatibilités booléennes voire préférentielles (WILKS)). Elles seules, en effet, permettront de valider l'une des trois relations de dépendance : à savoir celle qui s'établie entre "produits" et "algues" et non entre "vente" et "algues" ou "beauté" et "algues". Le traitement demandé ici est en tout état de cause syntaxico-sémantique.

• Dans la phrase (2) : la composante syntaxique seule suffit pour établir, sans ambiguïté, cette même relation de dépendance et par conséquent la RLS *(produits, SS_{aux}, algues)* .

Deux remarques importantes peuvent dès lors être faites, selon que l'on s'intéresse à l'aspect acquisition des RLS ou à leur utilisation :

Remarque 1 :

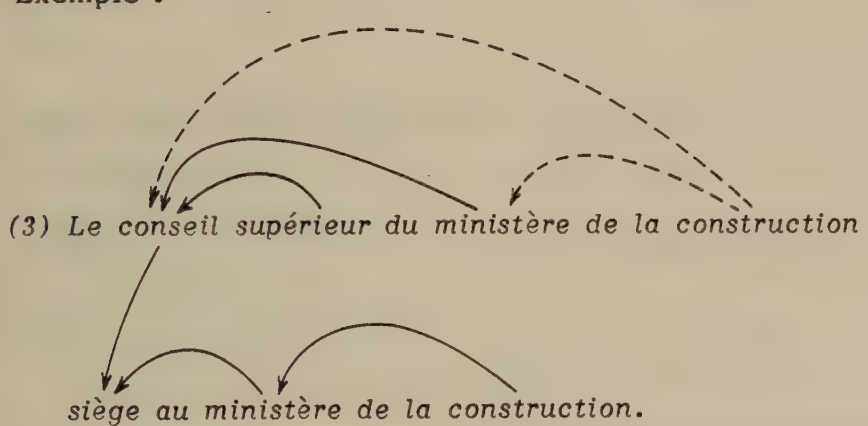
Certaines RLS peuvent être établies par la seule mise en oeuvre des informations grammatico-syntaxiques (e. g. *(produits, SS_{aux}, algues)* dans (2)).

Remarque 2 :

Dans un même corpus, une même RLS syntagmatique peut apparaître dans des structures syntaxiques différentes. Parmi celles-ci, certaines donnent lieu à des ambiguïtés de fléchage (e. g. phrase 1), d'autres n'en engendrent aucune (e. g. phrase 2). Dans ces conditions, il est possible, par l'acquisition et le cumul des RLS issues des unes (e. g. phrase 2), de lever les ambiguïtés qui surgissent dans les autres (e. g. phrase 1).

La redondance du discours, la synonymie tant lexicale que syntaxique (cf. ch. II, § IV), la répétition mais aussi les règles qui tendent à interdire ou à éviter cette répétition (pronoms, emploi du générique, etc.) contribuent à renforcer les deux remarques précédentes. Ces aspects, inhérents aux langues naturelles, accentuent en effet la coexistence de relations de dépendance "ambiguës" et "non-ambiguës" au sein d'un même corpus, voire d'un même texte, ou encore, comme le montre l'exemple suivant, d'une même phrase. (voir aussi fig. 69 (b)).

Exemple :

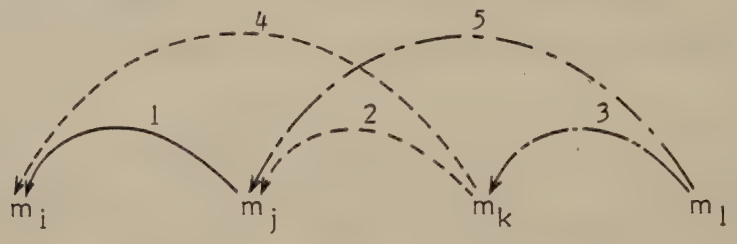


(Règlementation de l'urbanisme)

Ici, la même relation de dépendance qui existe entre "ministère" et "construction" est reconnue de façon "sûre" dans la deuxième chaîne nominale alors qu'elle reste "incertaine" dans la première chaîne nominale. La RLS *(ministère, SS_{de}, construction)* peut donc être acquise néanmoins et, lors d'un second balayage, permettre l'élimination de l'ambiguïté présente dans la première chaîne nominale.(Pour une discussion des limites de cette approche, voir conclusion, § 1).

Notons enfin l'aspect cumulatif de ce processus ou encore d'aide mutuelle qui peut s'instaurer entre composantes syntaxique et sémantique : l'application d'une RLS peut en effet être suivie de traitements syntaxiques particuliers qui, à leur tour, peuvent provoquer l'acquisition de nouvelles RLS et ainsi de suite.

A titre illustratif nous allons considérer la configuration littérale suivante, fléchée comme indiqué :



où l'on est en présence de deux ambiguïtés : (2 ou 4 ?) et (3 ou 5 ?).

Si l'on admet qu'en règle générale deux flèches ne peuvent se croiser, alors la validation de la dépendance 4 (ou respectivement 5) implique le rejet de la dépendance 5 (respectivement 4) et par conséquent la validation de la dépendance 3 (respectivement 2). La RLS correspondante peut dès lors être retenue.

Ce sont là les diverses remarques fondamentales qui sous-tendent le système que nous présentons et qui nous amènent à formuler l'hypothèse de base suivante:

> *Les RLS syntagmatiques, nécessaires pour la détermination des relations de dépendance, peuvent être obtenues par apprentissage au moyen de traitements syntaxico-sémantiques.*

Remarque teminologique :

Faisant précisément appel à ces RLS, l'analyse syntaxico-sémantique sera dite à apprentissage.

I. 2. Apprentissage des RLS syntagmatiques :

Le problème qui se pose maintenant est donc celui de la reconnaissance, parmi toutes les relations de dépendance qui ont été établies lors de l'étape précédente, de celles qui sont, dirons-nous, valides.

Toute relation de dépendance jugée comme telle est par suite retenue et la RLS correspondante apprise.

Cette validité ne peut cependant être décrétée d'une manière systématiquement sûre. A l'inverse, une relation de dépendance potentielle ne peut être non plus toujours rejetée avec certitude. Ainsi que nous le verrons, cela arrive notamment lorsqu'il y a lieu, pour effectuer cette reconnaissance, de faire intervenir les RLS (cf. ch. V, § IV).

La validité d'une dépendance ne peut dès lors être représentée par une simple variable à deux états.

Il est nécessaire de recourir à une évaluation "continue" (COULON & KAYSER 41) qui rendra possible les choix "préférentiels" (WILKS, 112) en cas d'ambiguïté.

Pour ce faire nous associons alors à chaque relation de dépendance une note censée mesurer son degré de validité et dont l'évaluation est fondée sur les divers calculs qui ont concouru à l'établir.

Nous verrons plus loin les différents critères qui président à l'évaluation de cette note. Nous retiendrons toutefois pour l'instant que cette note peut évoluer entre deux limites, notées **"V"** et **"D"** : **"V"** pour indiquer que la relation de dépendance établie est sûre (Valide) et que par conséquent la RLS correspondante peut être apprise ; **"D"** qu'elle est Douteuse et que par conséquent elle ne peut donner lieu à aucun apprentissage.

Nous voyons ainsi apparaître la nécessité d'un seuil d'apprentissage pour les relations de dépendance qui auront reçu une note comprise entre ces deux limites.

En effet, il n'est pas toujours nécessaire pour apprendre une RLS, que la relation de dépendance dont elle est issue soit notée "V". Inversement, il ne suffit pas toujours, en cas d'ambiguïté, à une relation de dépendance d'être la mieux classée pour donner lieu à l'apprentissage d'une RLS.

Plus généralement donc, une relation de dépendance se verra donner naissance à l'apprentissage d'une RLS si, étant la mieux classée, elle a reçu par ailleurs une note égale ou supérieure à un certain seuil fixé à l'avance. En particulier, il n'y aura aucune RLS apprise si ce seuil est fixé supérieur à "V".

La détermination de ce seuil ne peut toutefois être absolue. Notre but étant en définitive l'analyse des langues naturelles sous forme de dépendances, la détermination du seuil d'apprentissage sera empirique et sa valeur ajustée au vu des résultats obtenus.

Notons enfin que c'est également ce seuil qui permettra d'accélérer ou de ralentir l'apprentissage des RLS.

I. 3. RLS induites par la syntaxe :

Ainsi que nous l'avons vu, les RLS peuvent être obtenues soit par la seule mise en oeuvre de traitements syntaxiques, soit par la mise en oeuvre conjointe de traitements syntaxiques et de traitements sémantiques.

Nous appellerons RLS induites par la syntaxe les premières. Bien sûr, cette distinction ne constitue pas une caractéristique fondamentale si l'on a en vue seulement les RLS (puisque une même RLS peut, pour être obtenue, demander un traitement soit syntaxique soit syntaxico-sémantique selon les configurations dont elle est issue). Elle est par contre importante si l'on a en vue l'évaluation de la note qui doit être affectée aux relations de dépendance dont ces RLS découlent, ou encore l'apprentissage proprement dit de ces RLS.

En effet, les relations de dépendance dont la détermination ne réclame pas de contrôle sémantique sont (par définition) non-ambiguës : elles peuvent par conséquent recevoir, à l'issue de l'étape syntaxique, la note de validité la plus forte, soit "V"(*). Cette note sera évidemment définitive. Au contraire, les relations de dépendance réclamant un contrôle sémantique sont (par définition) ambiguës : elles peuvent alors recevoir, à l'issue de l'étape syntaxique, la note de validité la plus faible, soit "D"(*). Celle-ci pourra cependant être provisoire cette fois puisque les traitements ultérieurs auront pour but soit de la maintenir soit de l'améliorer (cf. ch. V, § IV. 2.).

Nous sommes donc désormais en mesure, si le seuil d'apprentissage retenu est inférieur à "V"(mais supérieur à "D"), d'initialiser l'apprentissage sémantique en enregistrant les RLS correspondant aux relations de dépendance notées "V" à l'issue des traitements syntaxiques(*).

Outre que les RLS induites par la syntaxe permettent ainsi d'initialiser le processus d'apprentissage, nous retiendrons également qu'elles sont les plus nombreuses (cf. ch. V, § I. 1.). Les traitements syntaxiques permettent en effet dans la majeure partie des cas, d'obtenir les relations de dépendance sans ambiguïté. C'est ce que nous pouvions déjà remarquer au travers des nombreux exemples illustratifs du chapitre III (cf. ch. III, fig. 40 et 69).

A titre indicatif nous donnons dans le tableau suivant les taux respectifs des relations de dépendance ayant été décrétées, à l'issue des traitements syntaxiques, valides (notées "V") ou douteuse (notées "D"). Ces taux ont été relevés sur trois textes notés T_1, T_2 et T_3 comptant respectivement 4700, 4650 et 4550 mots.

(*) Voir les figures 27, 29, 34, 35, 36, 38, 40, 46, 50, 54, 58, 61, 64, 66 et 69 du chapitre III.

	T_1	T_2	T_3	Total
Nbre de relations jugées valides	1037	1043	846	2926
Taux en %	89.47	88.02	91.26	89.45
Nbre de relations douteuses	122	142	81	345
Taux en %	10.53	11.98	8.74	10.55

fig. 87

Il convient de noter cependant que les RLS, induites par la syntaxe, ainsi apprises pourront présenter quelques erreurs. En effet, les options que nous avons adoptées lors des traitements syntaxiques (cf. ch. III, 2^e partie) ont laissé transparaître des cas de bruits totaux. Ces cas, ainsi que nous l'avons vu, correspondaient à des situations où parmi les relations de dépendance potentielles établies ne figuraient pas la ou les relations de dépendance finalement convoitées.

Dans ces conditions, toute relation de dépendance retenue, car non-ambiguë, ne peut qu'être erronée et la RLS apprise correspondante fausse(*).

Ces cas resteront cependant relativement rares puisque jamais les taux de bruit total n'ont excédé 10 % des relations établies (cf. ch. III).

A titre indicatif, nous donnons dans le tableau suivant, le taux des relations de dépendance erronées ayant reçu néanmoins la note "V" en raison de leur non-ambiguïté. Les comptages ont été effectués sur un extrait du texte T_1, d'environ 3 800 mots.

(*) En fait une RLS peut être "acceptable" bien qu'issue d'une relation de dépendance erronée. Ces cas sont cependant aléatoires et ne peuvent dès lors être retenus pour la mesure des performances du système.

Relations ayant reçu la note :	V		D
	avec succès	sans succès	
Nbre absolu des relations	883	30	98
Taux en % rapporté au nbre total de rela. notées V	96.7	3.3	-
Taux en % rapporté au nbre total de rela. établies	87.3	3	9.7

fig. 88.

La figure 89 donne le détail des comptages consignés dans ce tableau relativement aux diverses relations de dépendance reconnues.

Deux remarques peuvent alors être faites :

- la première est relative aux relations de dépendance erronées ayant été jugées valides. Leur nombre est relativement faible. Nous les négligerons lorsque nous nous intéresserons à l'évolution de l'apprentissage sémantique ; d'autant plus qu'elles auront peu de chance d'être sollicitées (cf. ch. II, b, p. 32).
- la deuxième est relative aux relations de dépendance restées ambiguës. Elles représentent environ 10 % des relations établies et sont essentiellement le fait des relations de détermination (SS_p et SP).

II. Représentation des connaissances :

Par abus de langage, nous appellerons **"connaissances"** l'ensemble des RLS nécessaires à la résolution des ambiguïtés syntaxiques ou lexicales mises à jour lors des étapes précédentes.

Nom des rel. établies	SA			SP			SS_p			VS			VD			VI_p			VA		
Note attribuée aux rel. à l'issue de l'étape synt.	V		D	V		D	V		D	V		D	V		D	V		D	V		D
	avec succès	sans succès		avec succès	sans succès		avec succès	sans succès		avec succès	sans succès		avec succès	sans succès		avec succès	sans succès		avec succès	sans succès	
Nbre absolu des rel.	29	-	-	174	-	23	285	11	58	124	8	14	202	6	3	65	5	-	4	-	-
Taux en % rapporté aux nbres totaux resp. des diverses rel. notées V	100	-	-	100	-	-	96.3	3.7	-	94	6	-	97.1	2.9	-	92.9	7.1	-	100	-	-
Taux en % rapporté aux nbres totaux resp. des diverses rel. établies	100	-	-	88.3	-	11.7	80.5	3.1	16.4	84.9	5.5	9.6	95.8	2.8	1.4	92.9	7.1	-	100	-	-
Taux en % rapporté au nbre total des rel. notées V (ou resp. D)	3.18	-	-	19.06	-	23.47	31.21	1.2	59.18	13.58	0.88	14.29	22.12	0.66	3.06	7.12	0.55	-	0.44	-	-
Taux en % rapporté au nbre total des rel. établies	2.87	-	-	17.21	-	2.27	28.19	1.09	5.74	12.26	0.79	1.38	19.98	0.6	0.3	6.43	0.49	-	0.4	-	-

Longueur du texte: environ 3800 mots.

fig. 89 - Performances du traitement syntaxique

Ayant distingué deux types de relations, la représentation de cette connaissance devra donc maintenir cette distinction. Elle se fera par conséquent selon les deux axes classiques de la linguistique que sont les axes paradigmatique et syntagmatique :

- le premier, vertical, est relatif aux relations qui existent entre les unités susceptibles de commuter (54) . Il décrit les "substitutions équivalentes" possibles (115) .
- le deuxième, horizontal, est relatif aux relations que contractent entre elles les unités lexicales dans la chaîne écrite. Il permet notamment de rendre compte de l'expression idiomatique (115) .

Une relation de dépendance R_s existant entre deux mots m_i et m_j d'une phrase donnée sera donc représentée comme suit :

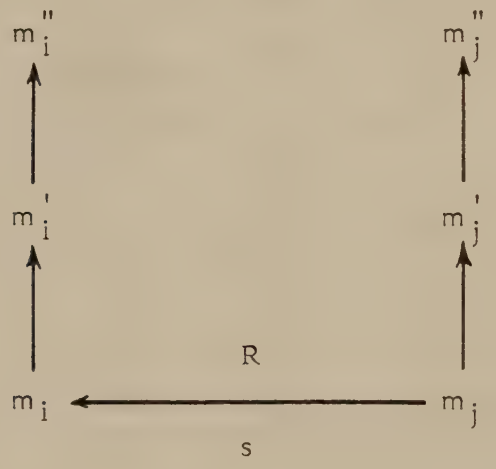


fig. 90.

- m'_i (resp. m'_j) étant le corrélatif lexical de m_i (resp. m_j) dans le premier niveau descriptif (cf. ch. IV, § X),
- m''_i (resp. m''_j) étant celui du deuxième niveau descriptif (cf. ch. IV, § X).

C'est cette structure qui constituera la **cellule descriptive** de base de ce que nous avons convenu d'appeler connaissance.

Exemple :

La relation de dépendance qui s'établit entre "affichages" et "muraux" dans la phrase suivante :

Les affichages muraux sont interdits

sera représentée comme suit :

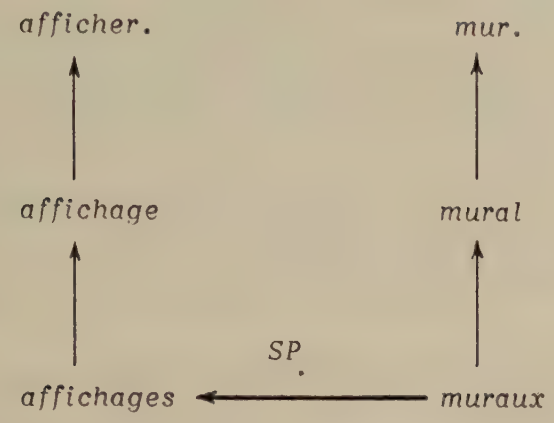


fig. 91

Plus généralement, la représentation d'une phrase sera obtenue par la connexion entre elles de cellules de ce type.

Ainsi, la phrase précédente recevra la représentation suivante :

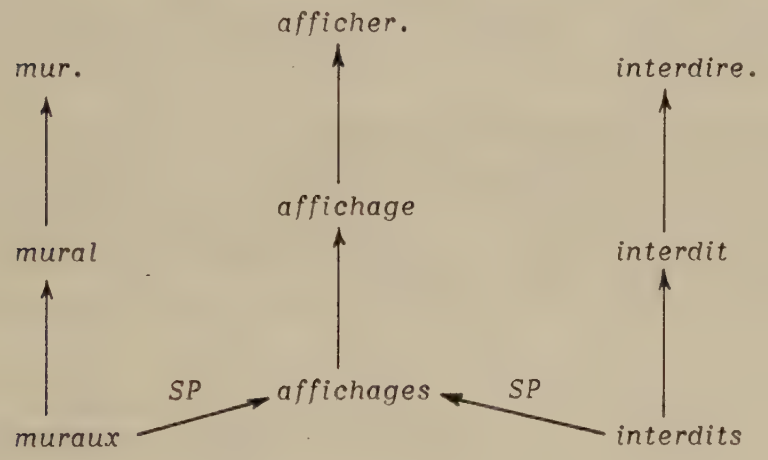


fig. 92.

Ainsi que nous allons le voir dans les prochains paragraphes, c'est cette **représentation à deux axes** , très nettement distingués, qui rendra possible une utilisation inférentielle de la connaissance. Cette représentation nous permettra en effet de simuler en partie, certaines règles transformationnelles et dès lors, de détecter certaines synonymies lexico-syntaxiques (paraphrases) (cf. ch. II, § III et IV). Nous verrons que ceci sera en particulier fondé sur la possibilité qu'ont les corrélatifs lexicaux d'hériter des relations qui s'établissent entre les éléments de la phrase.

Remarque :

La représentation à deux axes que nous avons adoptée n'est pas sans rappeler les représentations utilisées dans les réseaux sémantiques, où souvent l'on définit également deux axes, eux aussi fondés sur la nature des relations mises en jeu dans ces réseaux.

Ainsi, les deux axes que nous avons adoptés sont comparables à ceux qui sont dégagés par R.F. SIMMONS dans les réseaux sémantiques (107) .

R.J. BRACHMAN (cité par J.H. JAYEZ, 70) distingue également deux axes de représentation :

- l'un décrit les relations d'une classe à ses éléments ; on peut alors rapprocher cet axe de notre axe paradigmatique,
- l'autre concerne les liens qui s'établissent entre un concept et ses attributs descriptifs (une personne a une taille et un poids par exemple).

III. Mécanisme de résolution - Raisonnement par analogie :

Que l'on ait en vue la résolution des ambiguïtés syntaxiques ou des ambiguïtés lexicales, dans sa généralité, le problème que l'on se pose peut être formulé comme suit :

Etant donné la relation de dépendance potentielle (m_i, m_j, R_x), on se propose, compte tenu de la connaissance déjà accumulée, de déterminer le degré de validité que l'on peut accorder à cette relation (cf. fig. 93).

Pour cela, nous tenons le raisonnement suivant :

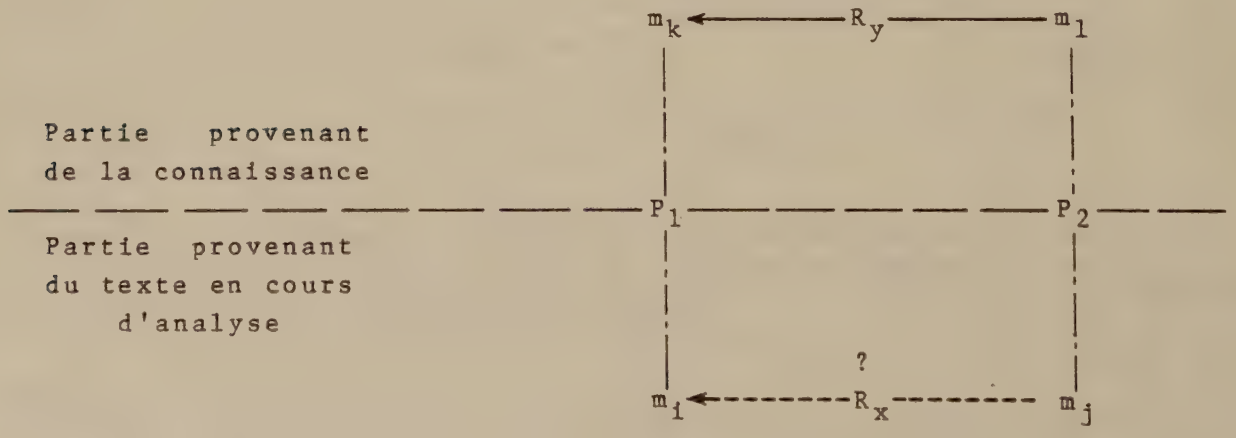


fig. 93.

Sachant qu'il existe :

- une relation de parenté sémantique P_1 (plus ou moins éloignée) entre m_i et m_k (axe paradigmatique),
- une relation de parenté sémantique P_2 entre m_j et m_l,
- une relation de dépendance R_y valide entre m_k et m_l (axe syntagmatique),

par analogie, nous déclarons alors que R_x est valide, son degré de validité étant d'autant plus élevé que :

- R_x et R_y sont proches (identiques ou non), et que
- P_1 et P_2 se rapprochent de l'identité.

La note de validité dépendra donc directement de P_1, P_2, R_x et R_y.

De plus, nous dirons que les deux triplets (m_i, m_j, R_x) et (m_k, m_l, R_y) sont équivalents.

Le processus ici mis en jeu, correspond donc, ainsi que nous pouvons le voir, à un raisonnement par analogie.

C'est précisément sur ce raisonnement qu'est fondée la résolution des ambiguïtés syntaxiques ou lexicales et, par suite, la composante sémantique du système.

Remarques :

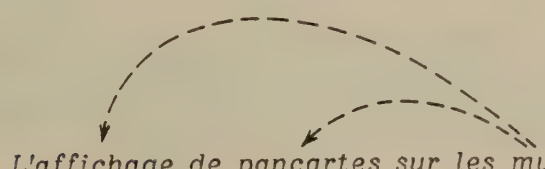
Ainsi que nous pouvons déjà le préssentir, c'est également ce raisonnement qui va nous permettre de détecter et d'établir les synonymies lexico-syntaxiques et, par conséquent de simuler certains mécanismes transformationnels. Dès lors, un apprentissage de ces transformations pourrait éventuellement être envisagé.

Rappelons à cet égard que la matrice des synonymies lexico-syntaxiques possibles donnée figure 3 n'est pas fournie au système, mais qu'elle pourrait être, éventuellement, établie automatiquement, par apprentissage.

IV. Résolution des ambiguïtés syntaxiques :

IV. 1. Mécanisme de résolution :

Il s'agit donc ici de résoudre les ambiguïtés qui auront été mises à jour lors de la construction des diverses relations de dépendance potentielles. C'est par exemple le cas dans la phrase suivante :



L'affichage de pancartes sur les murs est interdit.

où la liaison candidate entre "pancartes" et "murs" rivalise avec la liaison qui existe entre "affichage" et "murs".

Pour résoudre ces ambiguïtés, nous recourons donc au mécanisme précédemment décrit. A cet effet, nous supposerons, pour les besoins de l'exposé, que la cellule correspondant à la d´pendance *(affichages, muraux, SP)* de l'exemple précédent a été enregistrée dans la connaissance (cf. fig. 91).

Le schéma de résolution, pour l'exemple considéré, se présente donc comme suit :

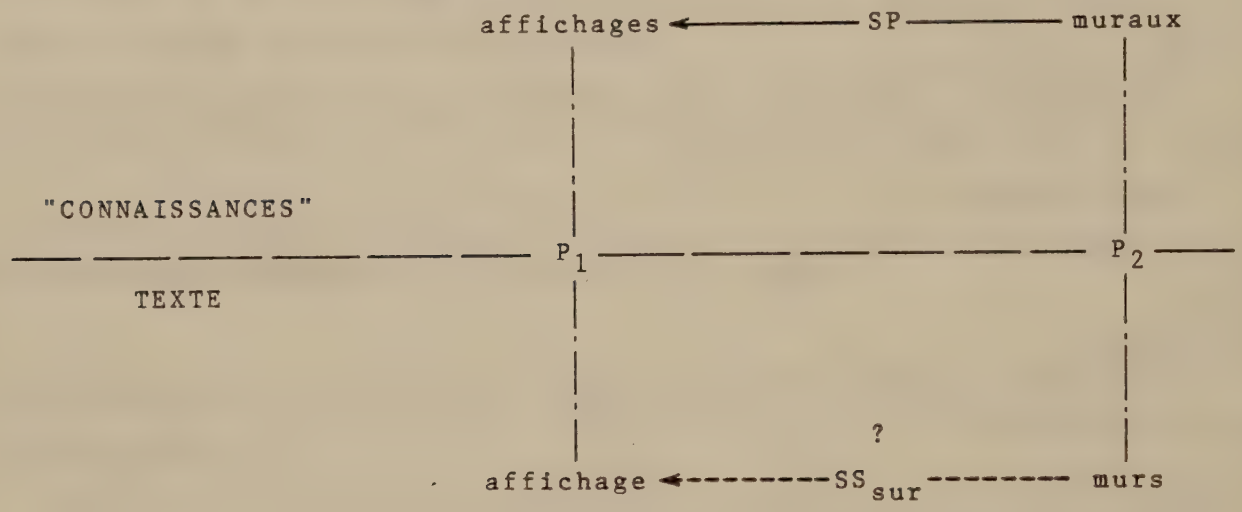


fig. 94.

la distance séparant les deux relations donnant la note de validité à affecter à la relation du texte ou de la phrase.

En représentant intégralement les cellules descriptives de l'une et de l'autre relation nous obtenons alors le schéma développé suivant (voir p. 239), où l'on remarque :

- d'une part, que les premiers éléments respectifs des deux relations sont identiques à partir du premier niveau descriptif de l'axe paradigmatique, et que les seconds le sont à partir du deuxième niveau,
- d'autre part, que les deux liaisons syntagmatiques mises en jeu sont différentes.

Les deux relations peuvent donc, en vertu du raisonnement par analogie précédemment décrit, être déclarées "équivalentes", mais il reste néanmoins à évaluer le degré de confiance que l'on peut attribuer à ce jugement.

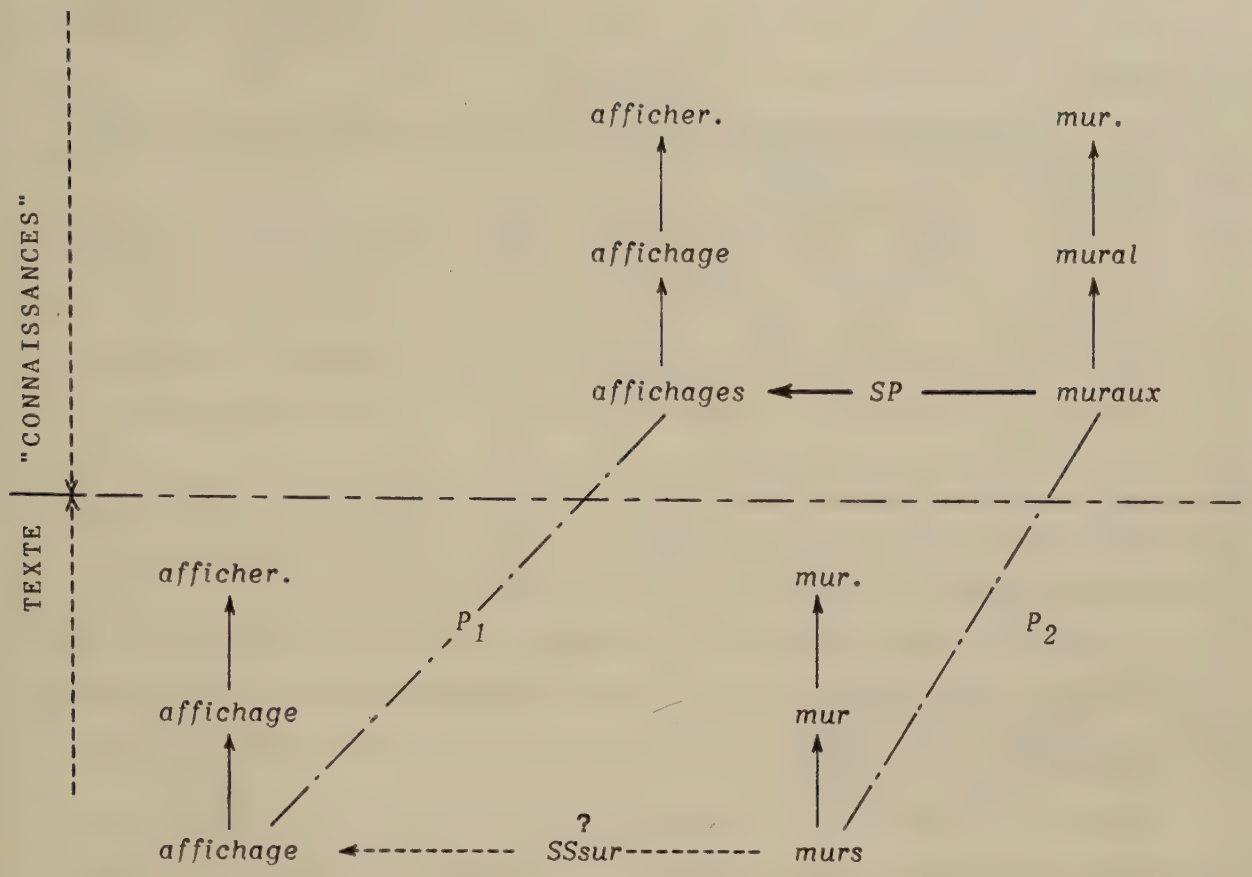


fig. 95.

IV. 2. Evaluation de la note de validité :

C'est le schéma développé de la figure 95 qui va nous permettre d'effectuer cette évaluation.

Nous pouvons remarquer en effet que les distances paradigmatiques peuvent prendre trois valeurs selon qu'il y a identité entre les éléments des relations elles-mêmes, entre les éléments des premiers niveaux descriptifs, ou enfin, entre les éléments des derniers niveaux descriptifs.

Par convention, nous associerons respectivement à ces trois cas les valeurs **0**, **1** et **2**.

La distance syntagmatique par contre ne prendra que deux valeurs (**0** ou **1**) selon que les deux liaisons syntagmatiques sont de même nature ou non.

En désignant par d_1 et d_2 les distances paradigmatiques associées respectivement à P_1 et P_2 et par d_3 la distance syntagmatique, il est alors possible d'établir le tableau de notation donné figure 96.

Remarque :

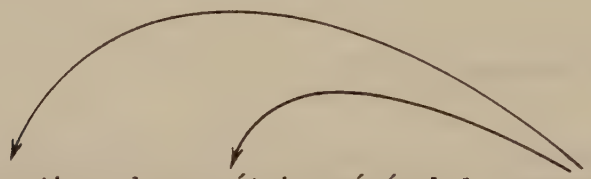
En fait, nous pourrions choisir pour ces distances des domaines de définition plus vastes :

- pour d_3 en établissant, soit arbitrairement, soit automatiquement (fréquence par exemple), l'ensemble des diverses distances inter-relations. C'est ici qu'interviendrait la matrice des synonymies lexico-syntaxiques (cf. fig. 3). Elle contiendrait alors non pas 0 ou 1 mais un poids reflétant, par exemple, la fréquence avec laquelle se produisent les diverses synonymies lexico-syntaxiques possibles.
- pour d_1 et d_2 en faisant intervenir les relations satellites mises en jeu, la négation notamment. Il serait également possible de faire intervenir des mécanismes de calcul de distances plus élaborés encore (96) .

Les notes de validité choisies et consignées dans ce tableau restent évidemment arbitraires ; mais elles sont néanmoins réajustées selon la qualité des résultats obtenus. En particulier, nous n'avons, jusqu'ici, relevé aucun mauvais choix consécutif à cette notation.

Il convient cependant de remarquer que malgré les critères de notation ainsi retenus ("proximités paradigmatiques et syntagmatiques"), certaines ambiguïtés peuvent persister. Plusieurs liaisons qui rivalisent peuvent en effet recevoir une même note à l'issue de la consultation de la connaissance.

Ce serait par exemple le cas dans la phrase suivante :



Les déclarations du secrétaire général du gouvernement

où les deux relations (déclarations, gouvernement, SS_{du}) et (secrétaire, gouvernement, SS_{du}) qui rivalisent sont toutes deux vraisemblables, et qui peuvent dès lors avoir été enregistrées dans la connaissance.

Dans de tels cas nous avons décidé de faire intervenir, en sus, un autre critère de choix, fondé cette fois sur la "proximité géographique" des termes des relations candidates.

Nous avons en effet émis l'hypothèse suivante :

> *Si plusieurs relations de dépendance ont, à l'issue de la consusltation de la connaissance, reçu une même note de validité, alors c'est la relation de dépendance dont les termes sont les plus proches géographiquement qui est retenue.*

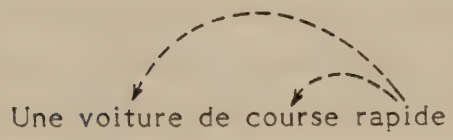
d_3	d_2	d_1	Note de validité	commentaires et exemples
0	0	0	U	Les cellules sont identiques
0	0	1	T	
0	0	2	Q	
0	1	0	T	
0	1	1	S	
0	1	2	P	
0	2	0	Q	
0	2	1	P	
0	2	2	N	réparation d'un immeuble/ réfection d'un édifice
1	0	0	R	hériter quelque chose/hériter de quelque chose
1	0	1	Q	
1	0	2	N	
1	1	0	Q	
1	1	1	P	
1	1	2	H	
1	2	0	I	
1	2	1	H	
1	2	2	F	réparer un immeuble/réfection d'un édifice

fig. 96. Table de notation.

Ces notes, établies empiriquement, doivent être considérées comme étant dans le voisinage de "V".

Dans la phrase précédente c'est donc la relation (secrétaire, gouvernement, SS_{du}) qui sera retenue.

Cette hypothèse n'est cependant pas toujours vérifiée. Ainsi dans :



c'est la relation (voiture, rapide, SP) qui est à retenir et non la relation (course, rapide, SP).

Nous verrons néanmoins que rarement cette règle a été mise en défaut (cf. ch. V, § IV. 4. 2.).

A cet égard il convient de remarquer qu'un autre critère aurait pu également être retenu. En effet, la fréquence des RLS stockées pourrait elle aussi apporter des indications permettant d'effectuer ou d'améliorer le choix en cas d'ambiguïté persistante. Ce serait là un moyen de tenir compte du contexte textuel dans lequel est plongée la phrase à analyser (cf. ch. I, § I).

IV. 3. Apprentissage. Organigramme général du traitement syntaxico-sémantique :

Nous allons dans ce paragraphe finir d'exposer le mécanisme d'apprentissage de l'analyse syntaxico-sémantique.

Notons tout d'abord que les relations sélectionnées par l'apport de la connaissance peuvent à leur tour être enregistrées, si évidemment la note qu'elles ont reçu est supérieure au seuil d'apprentissage retenu. Elles viendront ainsi s'ajouter aux relations induites par la syntaxe. Par rapport à ces dernières leur nombre restera cependant relativement faible. (cf. fig. 99)

La figure 97 donne l'organigramme désormais complet de l'analyse syntaxico-sémantique. On y remarque que les traitements syntaxiques et sémantiques sont imbriqués puisque l'appel à la connaissance se fait au fur et à mesure que l'on rencontre des ambiguïtés.

D'une façon plus générale, l'analyse d'une phrase ou d'un texte se fera en plusieurs passages. En effet, lors de l'analyse d'un texte il peut subsister certaines ambiguïtés syntaxiques en raison de la défaillance de certaines RLS syntagmatiques. Or celles-ci peuvent être acquises lors de traitement ultérieurs (au niveau d'un texte, c'est par exemple lors du traitement de phrases de rang supérieur).

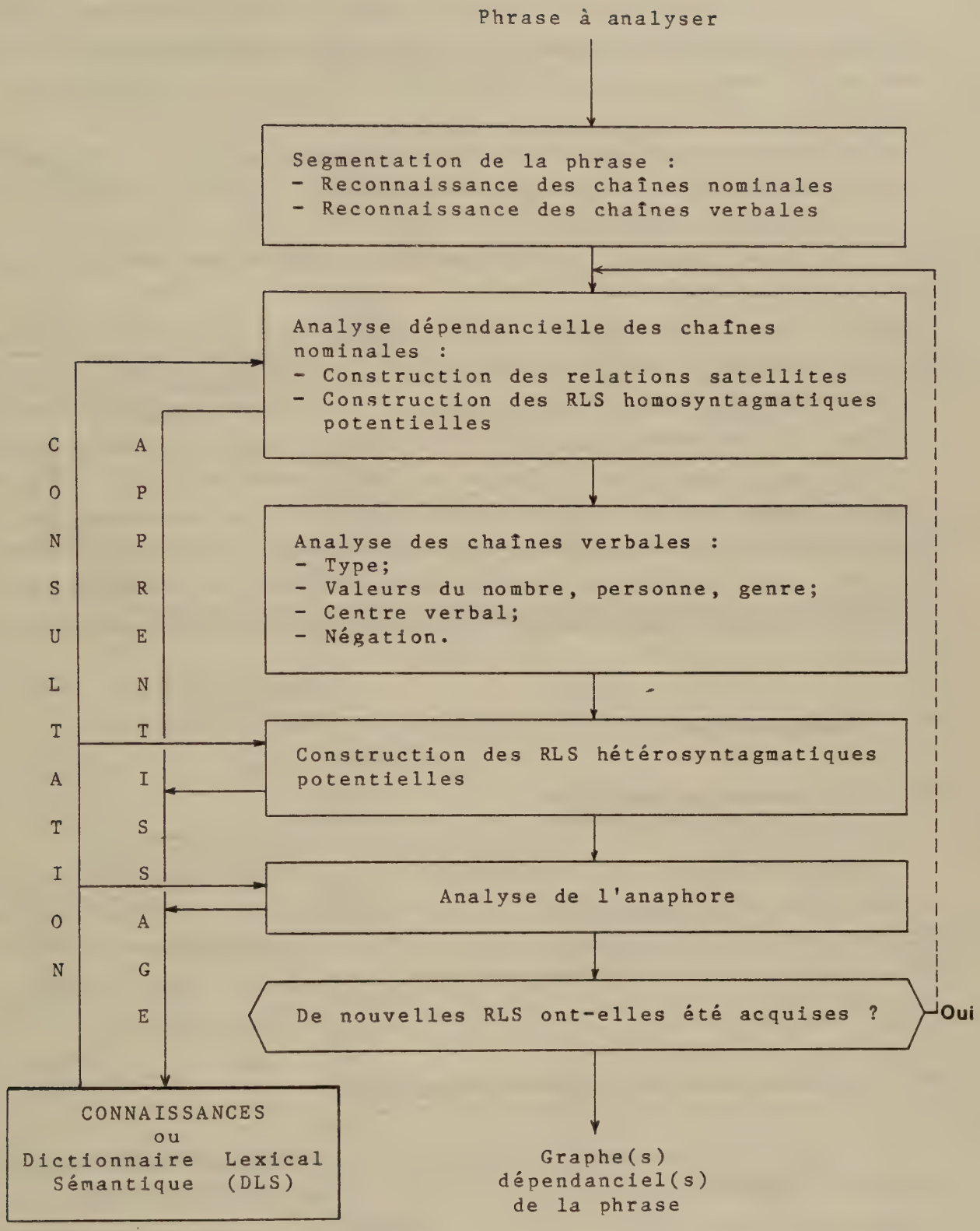


fig. 97. Organigramme général du traitement syntaxico-sémantique.

Il y a donc lieu d'effectuer des "va et vient" ou, plus exactement, de rebalayer les ambiguïtés persistantes, si toutefois de nouvelles RLS ont été depuis acquises (cf. retour en pointillé sur la fig. 97.).

Remarques :

Les expériences que nous avons menées (14 000 mots environ ont été analysés) n'ont pu nous permettre d'observer la convergence de l'apprentissage en terme de performance de l'algorithme (cf. § IV. 4. 3.).

A cet égard cependant, nous conjecturons que l'on pourrait distinguer deux types de convergence :

- l'une est locale, c'est à dire portant sur l'analyse de textes de même nature. Nous entendons par là que les textes à analyser appartiennent à un même champ ou discipline que les textes à partir desquels a été effectué l'apprentissage sémantique. La convergence dans ce cas devrait être alors relativement rapide.
- l'autre est globale, c'est à dire portant sur l'analyse de textes d'origines diverses. L'apprentissage est donc effectué à partir de textes composites. La convergence serait alors plus lente, ou même fortement compromise (KAYSER, 74). En effet, une RLS vraie dans un domaine peut être fausse dans un autre ; par exemple (corps, SP, fini) est valide en mathématiques mais non en anatomie.

IV. 4. Expérimentation :

IV. 4. 1. Exemples de résolutions :

Nous donnons dans les figures qui suivent quelques cas d'ambiguïtés syntaxiques ayant été résolues grâce à un contrôle sémantique effectué à partir de RLS syntagmatiques acquises automatiquement. Ce sont les dépendances marquées de trois étoiles (cf. fig. 98) qui indiquent que la "connaissance" est intervenue.

IV. 4. 2. Performances du traitement sémantique :

Lors des contrôles sémantiques, diverses situations génératrices d'erreurs peuvent se présenter selon que les RLS dont on a besoin figurent totalement ou en partie dans la connaissance.

En effet, lorsque plusieurs relations de dépendance rivalisent d'une part, et que, d'autre part, elles sont toutes vraisemblables, deux types d'erreurs peuvent alors se présenter, outre les cas de silences partiels vus dans le chapitre III :

```
        PHRASE NO : 13   ELLE COMMENCE AU MOT NO:  1631

        $00000110 LES MEMBRES DU CONSEIL SUP#ERIEUR DU MINIST∂ERE DE LA CONSTRUCTION SONT NOMM#ES POUR
        DEUX ANS PAR ARR-ET#E DU MINISTRE DE LA CONSTRUCTION .

*******  CHAINES NOMINALES RECONNUES

        LES MEMBRES DU CONSEIL SUP#ERIEUR DU MINIST∂ERE DE LA CONSTRUCTION

  1633  1                  LES       MEMBRES
  1635  2  SS   DU                   CONSEIL                 1 V
  1636  3  SP                        SUP#ERIEUR              2 V
  1638  4  SS   DU                   MINIST∂ERE              2 T ***    1 D
  1641  5  SS   DE         LA        CONSTRUCTION            4 T ***    2 D   1 D
-------------------------------------------------------------------------------------------------------
        POUR DEUX ANS PAR ARR-ET#E DU MINISTRE DE LA CONSTRUCTION

  1645  1        POUR               DEUX ANS
  1647  2                           ARR-ET#E
  1649  3  SS   DU                   MINISTRE                2 V
  1652  4  SS   DE         LA        CONSTRUCTION            3 T ***    2 D
-------------------------------------------------------------------------------------------------------

*******  CHAINES VERBALES RECONNUES

        SONT NOMM#ES                                                              P NOMM#ES

  1633  1                  LES       MEMBRES
  1643  5  SP                        NOMM#ES                 1 V

  1643  5                           NOMM#ES
  1645  6  VC   POUR                DEUX ANS                5 V
  1647  7  VS                        ARR-ET#E                5 V
-------------------------------------------------------------------------------------------------------

        PHRASE NO : 15   ELLE COMMENCE AU MOT NO:  5112

        L' INTERPOSITION ∂A CHAQUE #ETAGE D' UNE COUPURE ∂A L' AIR LIBRE ENTRE LA PORTE D' ACC∂ES ∂A L'
        ESCALIER ET LES PORTES DES LOGEMENTS DISPENSE D' AM#ENAGER DEUX MOYENS D' #EVACUATION DISTINCTS .

.......................................................................................................

*******  CHAINES NOMINALES RECONNUES

        L' INTERPOSITION ∂A CHAQUE #ETAGE D' UNE COUPURE ∂A L' AIR LIBRE

  5113  1                  L'        INTERPOSITION
  5116  2  SS   ∂A         CHAQUE    #ETAGE                  1 V
  5119  3  SS   D'         UNE       COUPURE                 2 D   1 D
  5122  4  SS   ∂A         L'        AIR                     3 D   2 D   1 D
  5123  5  SP                        LIBRE                   4 T ***    3 D   2 D   1 D
-------------------------------------------------------------------------------------------------------
        ENTRE LA PORTE D' ACC∂ES ∂A L' ESCALIER

  5126  1        ENTRE     LA        PORTE
  5128  2  SS   D'                   ACC∂ES                  1 V
  5131  3  SS   ∂A         L'        ESCALIER                2 D   1 D
-------------------------------------------------------------------------------------------------------
        LES PORTES DES LOGEMENTS

  5134  1                  LES       PORTES
  5136  2  SS   DES                  LOGEMENTS               1 V
-------------------------------------------------------------------------------------------------------
        DEUX MOYENS D' #EVACUATION DISTINCTS

  5141  1                  DEUX      MOYENS
  5143  2  SS   D'                   #EVACUATION             1 V
  5144  3  SP                        DISTINCTS               1 V
-------------------------------------------------------------------------------------------------------
.......................................................................................................

*******  CHAINES VERBALES RECONNUES

        DISPENSE D' AM#ENAGER                                                     S AM#ENAGER

  5113  1  VS                L'        INTERPOSITION          10 V
  5139 10                    '         AM#ENAGER

  5139 10  SP                          AM#ENAGER              11 V
  5141 11                    DEUX      MOYENS
-------------------------------------------------------------------------------------------------------
```

fig. 98. (a) Comparer, pour la première phrase, avec la fig. 69. (b).

```
        PHRASE NO :115   ELLE COMMENCE AU MOT NO:  9010

        $00001930 LE MULTIPLE DU SALAIRE MINIMUM INTERPROFESSIONNEL GARANTI PR#EVU ∂A L' ARTICLE 26 DU
        D#ECRET DU 24 D#ECEMBRE 1963 EST FIX#E ∂A 400 POUR CHAQUE UNIT#E COMPOSANT LE FOYER .

.....

*******  CHAINES NOMINALES RECONNUES

        LE MULTIPLE DU SALAIRE MINIMUM INTERPROFESSIONNEL GARANTI PR#EVU ∂A L' ARTICLE 26 DU D#ECRET

  9012  1                    LE        MULTIPLE
  9014  2  SS   DU                     SALAIRE                 1 V
  9015  3  SP                          MINIMUM                 2 D    1 D
  9016  4  SP                          INTERPROFESSIONNEL      2 D    1 D
  9017  5  SP                          GARANTI                 2 D    1 D
  9018  6                             PR#EVU
  9021  7  VC   ∂A        L'          ARTICLE                 5 D    6 T ***
  9022  8                             26
  9024  9  SS   DU                     D#ECRET                 7 V
-------------------------------------------------------------------------------------------------------
        POUR CHAQUE UNIT#E
-------------------------------------------------------------------------------------------------------
        LE FOYER
-------------------------------------------------------------------------------------------------------
.....

*******  CHAINES VERBALES RECONNUES

        EST FIX#E                                                          P FIX#E

  9012  1                    LE        MULTIPLE
  9030  5  SP                          FIX#E                   1 V                                C. D'AGENT

  9030  5                             FIX#E
  9035  6  VC   POUR        CHAQUE    UNIT#E                  5 V
-------------------------------------------------------------------------------------------------------
        COMPOSANT                                                          S COMPOSANT

  9035  5  VS   POUR        CHAQUE    UNIT#E                  6 V
  9036  6                             COMPOSANT

  9036  6  SP                          COMPOSANT               7 V
  9038  7                    LE        FOYER
-------------------------------------------------------------------------------------------------------


        PHRASE NO : 72   ELLE COMMENCE AU MOT NO: 12481

        $&F&B&J3&A1110&A&H&A LA &T.&V.&A. S' APPLIQUE ∂A TOUTES LES OP#ERATIONS QUI REL∂EVENT D' UNE
        ACTIVIT#E INDUSTRIELLE OU COMMERCIALE QUELLE QUE SOIT LA FORME OU LA NATURE DE L' INTERVENTION DE LA
        PERSONNE QUI EFFECTUE CES OP#ERATIONS .

        LA NATURE DE L' INTERVENTION DE LA PERSONNE

 12504  1                    LA        NATURE
 12507  2  SS   DE           L'        INTERVENTION            1 V
 12510  3  SS   DE           LA        PERSONNE                2 F ***    1 D
-------------------------------------------------------------------------------------------------------
        EFFECTUE                                                           S EFFECTUE

 12504  5  VS                LA        NATURE                  8 D
 12507  6  VS   DE           L'        INTERVENTION            8 D
 12510  7  VS   DE           LA        PERSONNE                8 T ***
 12512  8                             EFFECTUE

 12512  8  SP                          EFFECTUE                9 V
 12514  9                    CES       OP#ERATIONS
-------------------------------------------------------------------------------------------------------
```

fig. 98. (b)

PHRASE NO :123 ELLE COMMENCE AU MOT NO: 9330

$00001950 LE CANDIDAT @A L' ACCESSION @A LA PROPRI#ET#E POURRA JUSTIFIER DE SES RESSOURCES PAR TOUS MOYENS DE PREUVE , NOTAMMENT PAR LA PRODUCTION DE L' AVERTISSEMENT D#ELIVR#E PAR LE DIRECTEUR DES CONTRIBUTIONS DIRECTES POUR L' ACQUIT DE L' IMP¬OT SUR LE REVENU DES PERSONNES PHYSIQUES .

******* CHAINES NOMINALES RECONNUES

LE CANDIDAT @A L' ACCESSION @A LA PROPRI#ET#E

9332	1			LE	CANDIDAT			
9335	2	SS	@A	L'	ACCESSION	1 V		
9338	3	SS	@A	LA	PROPRI#ET#E	2 T ***	1 D	

DE SES RESSOURCES PAR TOUS MOYENS DE PREUVE

9343	1		DE	SES	RESSOURCES			
9346	2			TOUS	MOYENS			
9348	3	SS	DE		PREUVE	2 V		

PAR LA PRODUCTION DE L' AVERTISSEMENT D#ELIVR#E PAR LE DIRECTEUR DES CONTRIBUTIONS DIRECTES

9353	1			LA	PRODUCTION			
9356	2	SS	DE	L'	AVERTISSEMENT	1 V		
9357	3	SP			D#ELIVR#E	2 V		
9360	4	VS		LE	DIRECTEUR	3 V		
9362	5	SS	DES		CONTRIBUTIONS	4 V		
9363	6	SP			DIRECTES	5 V		

POUR L' ACQUIT DE L' IMP¬OT SUR LE REVENU DES PERSONNES PHYSIQUES

9366	1		POUR	L'	ACQUIT			
9369	2	SS	DE	L'	IMP¬OT	1 V		
9372	3		SUR	LE	REVENU			
9374	4	SS	DES		PERSONNES	3 V		
9375	5	SP			PHYSIQUES	4 V		

******* CHAINES VERBALES RECONNUES

POURRA JUSTIFIER S JUSTIFIER

9332	1	VS		LE	CANDIDAT	4 V		
9340	4				JUSTIFIER			COMPLEMENT ?

SUR LES POINTS MENTIONN#ES AU PARAGRAPHE &I DE L' ARTICLE 1ER DU D#ECRET N. 61-1036

6191	1		SUR	LES	POINTS			
6192	2	SP			MENTIONN#ES	1 V		
6194	3	VC	AU		PARAGRAPHE	2 V		
6195	4				&I			
6198	5	SS	DE	L'	ARTICLE	3 V		
6199	6	SP			1ER	5 T ***	3 D	
6201	7	SS	DU		D#ECRET	5 T ***	3 D	
6202	8	SE			N.	7 V		
6203	9	SE			61-1036	7 V		

AVEC LE CONCOURS DU &CR#EDIT FONCIER DE &FRANCE

7700	1		AVEC	LE	CONCOURS			
7702	2	SS	DU		&CR#EDIT	1 V		
7703	3	SP			FONCIER	2 T ***	1 D	
7705	4	SS	DE		&FRANCE	2 V		

---

fig. 98. (c)

- le premier, est nous l'avons vu, consécutif au choix de la relation mettant en oeuvre les termes les plus proches géographiquement, lorsque toutes les relations candidates ont obtenu la même note de validité ;
- le deuxième est dû au manque d'apprentissage qui a pour conséquence le choix d'une relation de dépendance possible mais non souhaitée. C'est par exemple le cas dans la chaîne nominale suivante :

```
-------------------------------------------------------------------------------
        PAR DES OUVERTURES D' #EVACUATION DE L' AIR

3803  1                 DES       OUVERTURES
3805  2  SS   D'                  #EVACUATION              1 V
3808  3  SS   DE        L'        AIR                      2 D   1 F ***
-------------------------------------------------------------------------------
```

où la relation à retenir est en fait (évacuation, air, SS_{de}).

A titre indicatif, nous donnons dans le tableau suivant les taux d'erreur et de résolution relevés sur le même extrait d'environ 3 800 mots du texte T_1. Ce tableau reprend et complète le tableau de la figure 88.

Relations ayant reçu la note	V		D		
Analyse ou résolution	correcte	incor.	correcte	incor.	non résolues
Nbre absolu des relations	883	30	17	3	78
Taux en % rapporté aux nbres totaux respectifs des rela. ayant reçu la note V ou D	96.7	3.3	17.3	3.1	79.6
Taux en % rapporté au nbre total des relations établies	87.3	3.0	1.7	0.3	7.7

fig. 99.

Ainsi que nous pouvons le voir le taux d'erreur du traitement sémantique reste relativement faible. Ajoutons d'ailleurs que sur les trois erreurs relevées deux sont dues à un manque d'apprentissage (Les peformances données dans ce tableau correspondent à un apprentissage effectué sur les trois textes T_1, T_2 et T_3), la troisième étant due à un silence partiel dans le traitement syntaxique.

Les performances globales de l'algorithme d'analyse syntaxico-sémantique mesurées sur ce même texte se présentent donc comme suit :

Relations		
reconnues		non reconnues
correctement	incorrectement	
900	33	78
89 %	3.3 %	7.7 %

fig. 100.

Pour compléter ce tableau, il importe de noter que sur ces 78 cas un bon nombre pourrait être reconnu en faisant un apprentissage sur un univers plus vaste. Nous avons en effet comptabilisé 73 cas de ce genre , les 5 autres étant des silences partiels. (voir les figures 29, 34, 40, 50, 66, 69 (b), 69 (i), 98 (a), 98 (b)).

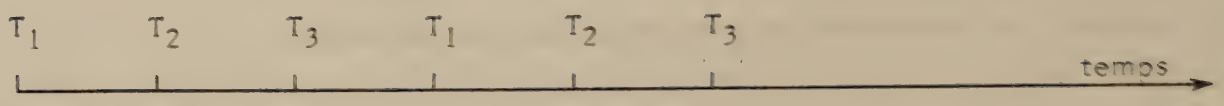
Remarque :

Pour réduire le nombre de ces ambiguïtés, nous pourrions là aussi faire intervenir la fréquence des relations restées douteuses. On peut en effet conjecturer qu'une relation de dépendance douteuse a d'autant plus de chance de s'avérer valide qu'elle est fréquente.

IV. 4. 2. Evolution de l'apprentissage sémantique :

Nous allons dans ce paragraphe donner l'évolution des performances de l'analyseur syntaxico-sémantique au cours du temps, l'expérimentation étant effectuée à l'aide des trois textes T_1, T_2 et T_3. Nous ne tiendrons pas compte cependant des taux d'erreur (syntaxique ou sémantique) qui sont, ainsi que nous l'avons vu, relativement faibles.

L'expérimentation que nous avons menée peut être décrite par le schéma suivant :



où, sur l'axe des temps, nous avons figuré l'ordre selon lequel les textes T_1, T_2 et T_3 ont été analysés ou réanalysés.

Nous avons alors obtenu les résultats exposés figure 101.

Notons qu'un troisième passage a été sans effet.

Sans pouvoir parler de la convergence de l'apprentissage en terme de taux de résolution (cf. ch. V, § IV. 3.), nous constatons néanmoins une amélioration sensible de la performance des traitements sémantiques au cours du temps.

V. Résolution de la polysémie :

Il s'agit de résoudre les ambiguïtés lexicales dont on a vu qu'elles étaient mises à jour lors de la description paradigmatique (cf. ch. IV, § X).

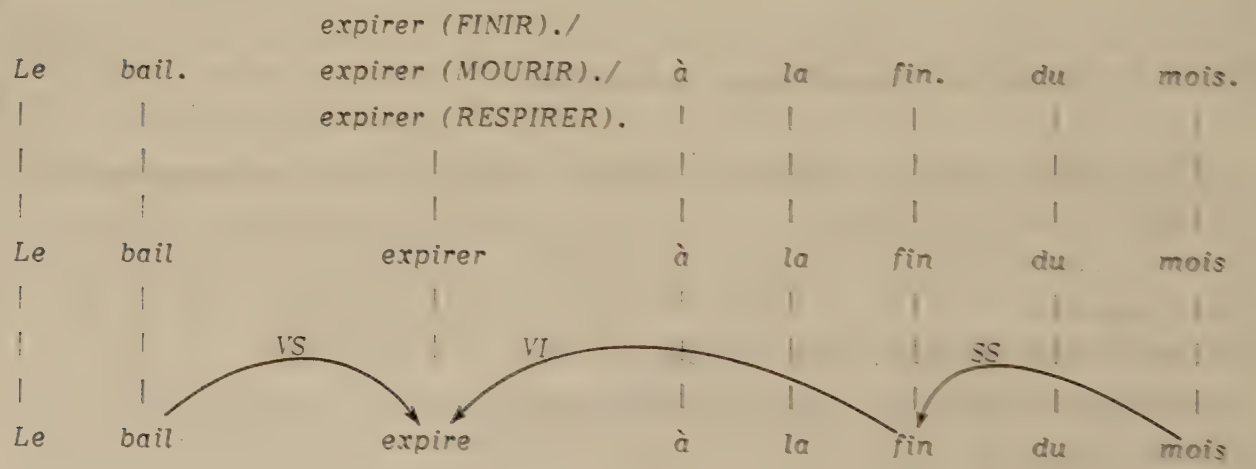
Seules par conséquent seront prises en considération les ambiguïtés qui ne peuvent être réduites au moyen d'informations grammaticales.

D'une façon plus précise, le but que l'on s'assigne est, lorsqu'un mot a reçu, lors de la description paradigmatique, plusieurs corrélatifs lexicaux, de déterminer parmi ceux-ci celui qui convient (ou ceux qui conviennent).

Reprenons, pour illustrer notre propos, la phrase :

Le bail expire à la fin du mois

dont la description à ce stade du traitement est :



T_1

HEURE : 18 H 49 MN DATE : 28 / 03 / 80
PERFORMANCES DE L'ALGORITHME

NBRE DE RELATIONS VALIDES =	1037	89.47 %
NBRE DE RELATIONS DOUTEUSES =	122	10.53 %
NBRE DE RELATIONS TOTAL =	1159	
NBRE DE RELATIONS RESOLUES =	19	1.64 %
APPORT DE LA CONNAISSANCE :	15.57 %	
PERFORMANCE GLOBALE DE L'ALGORITHME :	91.11 %	

HEURE : 21 H 25 MN DATE : 28 / 03 / 80
PERFORMANCES DE L'ALGORITHME

NBRE DE RELATIONS VALIDES =	1037	89.47 %
NBRE DE RELATIONS DOUTEUSES =	122	10.53 %
NBRE DE RELATIONS TOTAL =	1159	
NBRE DE RELATIONS RESOLUES =	23	1.98 %
APPORT DE LA CONNAISSANCE :	18.85 %	
PERFORMANCE GLOBALE DE L'ALGORITHME :	91.46 %	

T_2

HEURE : 19 H 39 MN DATE : 28 / 03 / 80
PERFORMANCES DE L'ALGORITHME

NBRE DE RELATIONS VALIDES =	1043	88.02 %
NBRE DE RELATIONS DOUTEUSES =	142	11.98 %
NBRE DE RELATIONS TOTAL =	1185	
NBRE DE RELATIONS RESOLUES =	39	3.29 %
APPORT DE LA CONNAISSANCE :	27.46 %	
PERFORMANCE GLOBALE DE L'ALGORITHME :	91.31 %	

HEURE : 21 H 42 MN DATE : 28 / 03 / 80
PERFORMANCES DE L'ALGORITHME

NBRE DE RELATIONS VALIDES =	1043	88.02 %
NBRE DE RELATIONS DOUTEUSES =	142	11.98 %
NBRE DE RELATIONS TOTAL =	1185	
NBRE DE RELATIONS RESOLUES =	61	5.15 %
APPORT DE LA CONNAISSANCE :	42.96 %	
PERFORMANCE GLOBALE DE L'ALGORITHME :	93.16 %	

T_3

HEURE : 21 H 00 MN DATE : 28 / 03 / 80
PERFORMANCES DE L'ALGORITHME

NBRE DE RELATIONS VALIDES =	846	91.26 %
NBRE DE RELATIONS DOUTEUSES =	81	8.74 %
NBRE DE RELATIONS TOTAL =	927	
NBRE DE RELATIONS RESOLUES =	13	1.40 %
APPORT DE LA CONNAISSANCE :	16.05 %	
PERFORMANCE GLOBALE DE L'ALGORITHME :	92.66 %	

HEURE : 22 H 09 MN DATE : 28 / 03 / 80
PERFORMANCES DE L'ALGORITHME

NBRE DE RELATIONS VALIDES =	846	91.26 %
NBRE DE RELATIONS DOUTEUSES =	81	8.74 %
NBRE DE RELATIONS TOTAL =	927	
NBRE DE RELATIONS RESOLUES =	26	2.80 %
APPORT DE LA CONNAISSANCE :	32.10 %	
PERFORMANCE GLOBALE DE L'ALGORITHME :	94.06 %	

fig. 101. Performance de l'algorithme au cours du temps.

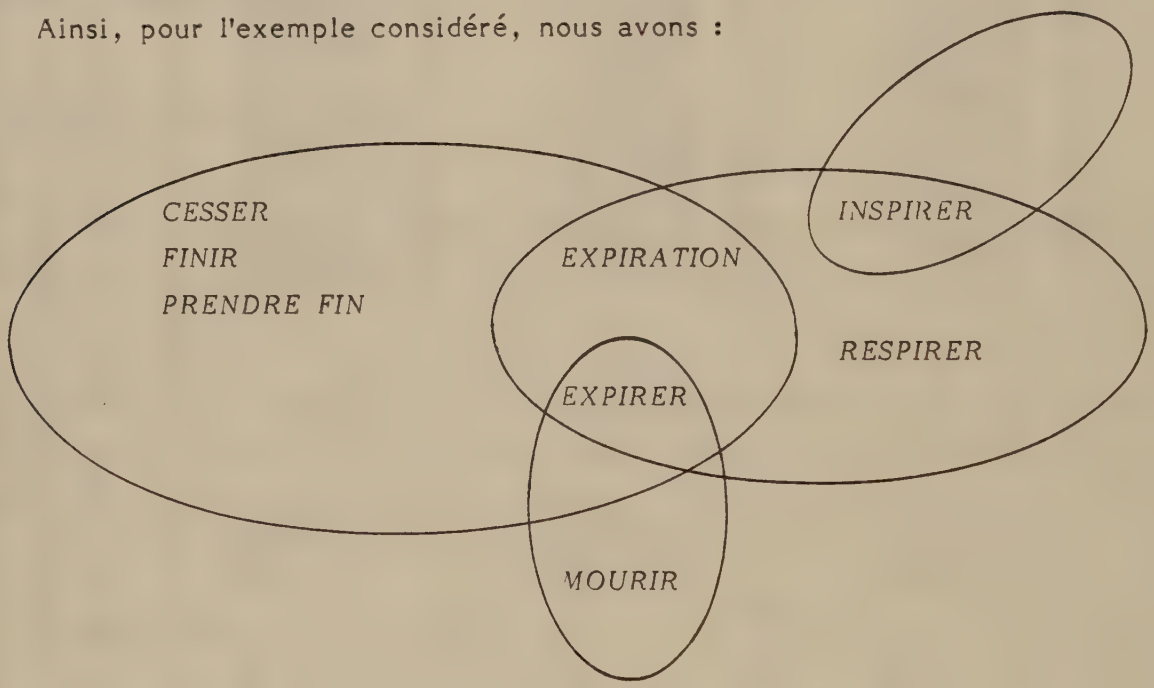
L'amélioration peut paraître faible mais cela ne doit pas faire illusion : comme nous l'avons vu plus haut, un apprentissage sur un univers plus vaste donnerait des résultats sans doute meilleurs. (cf. fig. 98 où les relations restées douteuses le sont du fait d'un manque d'apprentissage).

Il s'agit alors de sélectionner, pour le polysème "expire", le corrélatif lexical "expirer (FINIR)."

V. 1. Mécanisme de résolution :

Rappelons tout d'abord qu'un mot m est polysémique s'il se trouve à l'intersection de plusieurs familles (cf. ch. IV, § IV. 3. 3. 1.). Il est alors en relation paradigmatique avec les différents éléments des diverses familles auxquelles il appartient.

Ainsi, pour l'exemple considéré, nous avons :

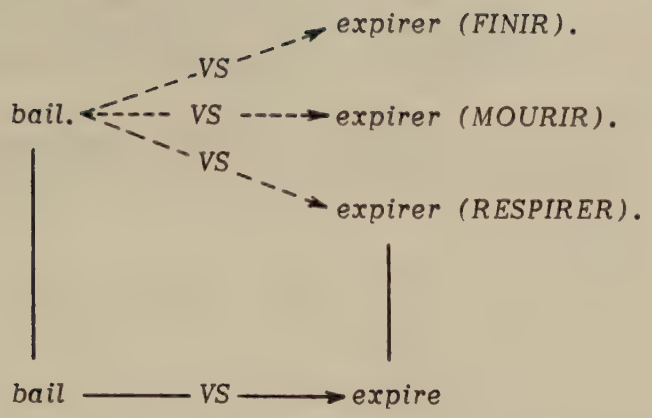


La résolution de la polysémie sera précisément fondée sur cette propriété.

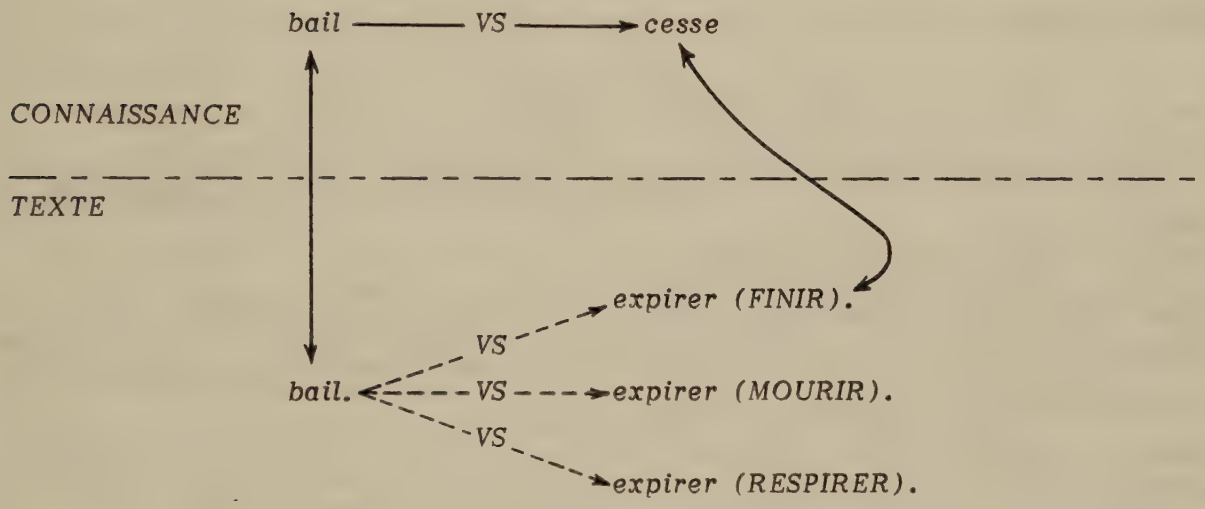
En effet, ce sont les relations syntagmatiques qu'auront pu contracter par ailleurs les éléments de ces familles avec l'autre mot de la relation (ici "bail") qui permettront d'effectuer un choix.

Ainsi, si la relation (bail, cesser, VS) a été préalablement enregistrée, le raisonnement par analogie précédemment décrit permettra alors de renforcer le choix portant sur le corrélatif "expirer (FINIR)."

Le problème revient donc, comme précédemment, à affecter une note de validité à chacune des relations héritées par les corrélatifs lexicaux respectifs des mots mis en relation dans la phrase. Ici à attribuer une note à chacune des relations héritées qui s'établissent entre les corrélatifs des mots "bail" et "expire" :



Le schéma de résolution pour cet exemple se présente donc comme suit :



Remarques :

Les relations verticales sont ici des relations d'appartenance. C'est la famille qui contient "cesse" qui sera retenue.

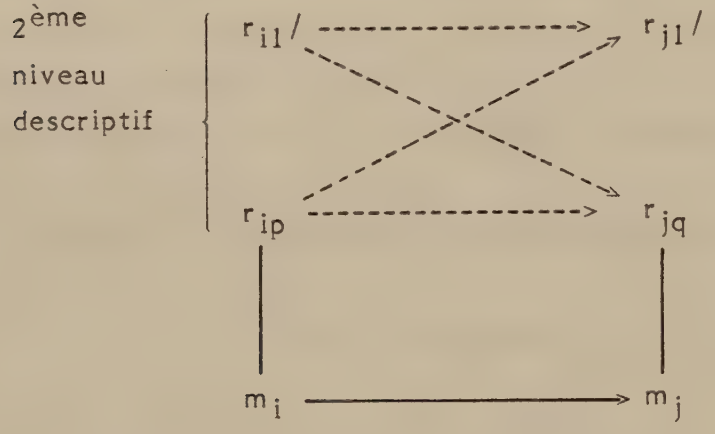
Il est clair cependant que ce mécanisme ne sera possible que si les familles considérées ne sont pas entièrement polysémiques, puisqu'il faudrait alors amorcer manuellement le processus en enregistrant préalablement dans la connaissance la relation qui convient.

V. 2. Evaluation de la note :

Nous allons considérer le cas le plus général où les deux mots, mis en relation dans la phrase, sont polysémiques.

Désignons par m_i et m_j ces deux mots et par $< r_{i1}, \ldots, r_{ip} >$ et $< r_{j1}, \ldots, r_{jq} >$ les ensembles de leurs corrélatifs respectifs du deuxième niveau descriptif.

Le problème consiste donc à affecter un poids à chacune des relations héritées qui s'établissent entre les éléments de ces deux ensembles.



La relation qui recevra le plus fort poids donnera alors les corrélatifs à retenir.

Pour ce faire nous opérons comme suit :

- dans une première étape, nous affectons, selon les mécanismes décrits au § IV, une note de validité à chacune des relations engendrées entre les éléments respectifs des familles associées aux mots m_i et m_j.
- dans une deuxième étape, nous procédons au calcul du poids à affecter à chacune des relations établies entre les familles respectives de m_i et m_j. Ce poids est obtenu en additionnant, pour chaque couple de familles, les notes précédemment attribuées aux relations qui s'établissent éventuellemnt entre leurs éléments.

Ces poids servent alors à ordonner les ensembles de familles respectivement associées à m_i et m_j.

V. 3. Expérimentation :

Nous donnons dans ce paragraphe juste quelques exemples de résolutions obtenues automatiquement.

Bien que nous ne puissions donner une mesure significative de la performance de l'algorithme, en raison du caractère limité de l'expérience, nous retiendrons néanmoins qu'au terme de l'apprentissage effectué à partir de T_1, T_2 et T_3, nous avons relevé, sur une vingtaine de cas traités, un cas de réso-

lution dont nous donnons le détail figure 102.

En outre, afin d'éprouver davantage l'algorithme, nous avons injecté dans la connaissance quelques RLS syntagmatiques supplémentaires. Toutes les ambiguïtés lexicales traitées ont été alors résolues sans erreur (cf. fig. 103).

Cette expérience reste cependant très limitée et il conviendra donc de la poursuivre.

VI. Conclusion :

Bien que les mesures portent sur un volume de textes modeste, les résultats obtenus semblent confirmer les hypothèses que nous avons avancées.

Nous soulignerons en particulier les résulats suivants :

- Environ 90 % des relations de dépendance établies sont non-ambiguës. Seulement 10 % réclament un contrôle sémantique.
- Sur ces 10 %, et pour un apprentissage sémantique effectué sur environ 14 000 mots, 30 % environ ont été résolues et on peut espérer améliorer ce résultat par un apprentissage portant sur plus de textes.

Pour ce qui est de la polysémie, nous retiendrons le caractère flou ou arbitraire qui entoure sa définition. A cet égard, nous partageons les analyses de COULON & AL. qui les conduisent à souligner le fait que *"suivant le niveau où l'on a décidé d'arrêter ses investigations, un même mot pourra apparaître doté d'un sens, de deux, ou de trente six"* (COULON & AL., 40, p.322).

Pourtant nous pensons que pour certaines applications, telles que la documentation automatique, nos approximations peuvent permettre un dégrossissage relativement satisfaisant.

```
                                                      RESOLUTION DE LA RELATION SUIVANTE:
                                                      ***********************************
          R∂EGLEMENTS                                                   EN VIGUEUR
   11  SP R#EGLEMENTER|R#EGLER(PAYER)                                  EN VIGUEUR


LES DIFFERENTS SENS
DU PREMIER MOT SONT:              LEUR FAMILLE EST:
.............................................................................................
R#EGLEMENTER                      R#EGLEMENTAIRE     R#EGLEMENTAIREMENT  R#EGLEMENTATION   R#EGLEMENTER
                                  R∂EGLEMENT
.............................................................................................
R#EGLER(PAYER)                    R#EGLER            R∂EGLEMENT
.............................................................................................


LES DIFFERENTS SENS
DU SECOND MOT SONT:               LEUR FAMILLE EST:
.............................................................................................
EN VIGUEUR                        EN VIGUEUR
.............................................................................................

IMPRESSIONS DES TABLEAUX:C  AVEC DES *
                         S1 AVEC DES +
                         S2 AVEC DES X

            EN VIGUEUR
```

	C	C	C	C	C	S1
R#EGLEMENTER	30	0	0	0	0	30
R#EGLER(PAYER)	15	0	0	0	0	15
	0	0	0	0	0	0
	0	0	0	0	0	0
	0	0	0	0	0	0

	S2	S2	S2	S2	S2
	45	0	0	0	0

```
APRES CONSULTATION DE CONNAISSANCE ON OBTIENT:
---------------------------------------------------------------------------------------------
          R∂EGLEMENTS                                                   EN VIGUEUR
   11  SP R#EGLEMENTER                                                  EN VIGUEUR
---------------------------------------------------------------------------------------------
```

Le tableau C donne les poids qu'ont reçu les relations héritées qui s'établissent entre les familles respectives de m_i (ici "règlements") et m_j (ici "en vigueur").

Remarque : Les poids établis ne sont jamais nuls car ils incluent au moins la note de la relation qui s'établie entre lesdits termes de la phrase.

Ci-après nous trouvons dans l'ordre :

- La phrase dont est issue la relation polysémique ci-dessus résolue ;
- La phrase dont est issue la RLS syntagmatique qui a permis cette résolution.

```
PHRASE NO : 95   ELLE COMMENCE AU MOT NO:  3199

$00000670 LES EXEMPTIONS DE PERMIS DE CONSTRUIRE R#ESULTANT DU PR#ESENT ARR-ET#E SONT ACCORD#EES
SANS PR#EJUDICE DE L' APPLICATION DES AUTRES LOIS ET R∂EGLEMENTS EN VIGUEUR .
.............................................................................................

PHRASE NO :102   ELLE COMMENCE AU MOT NO:  3404

CES CONSTRUCTIONS QUI FONT APPEL ∂A DES PRINCIPES NON TRADITIONNELS PR#ESENTENT SOIT DANS L'
AGENCEMENT INT#ERIEUR DES PLANS , SOIT DANS LES DIMENSIONS DE BASE DES LOCAUX , SOIT DANS L'
#EQUIPEMENT DES PI∂ECES PRINCIPALES OU SECONDAIRES , UN CARACT∂ERE NOVATEUR PAR RAPPORT AUX USAGES
ET ∂A LA R#EGLEMENTATION EN VIGUEUR .
```

fig. 102.

RESOLUTION DE LA RELATION SUIVANTE:
****

	COMPREND	LE	DOSSIER
VS	COMPR#EHENSION|COMPRENDRE(CONTENIR)	LE	DOSSIER

LES DIFFERENTS SENS DU PREMIER MOT SONT:

	LEUR FAMILLE EST:			
COMPR#EHENSION	COMPR#EHENSION	COMPRENDRE	COMPREHENSIBLE	COMPR#EHENSIF
COMPRENDRE(CONTENIR)	COMPRENDRE	CONTENIR		

LES DIFFERENTS SENS DU SECOND MOT SONT:

	LEUR FAMILLE EST:
DOSSIER	DOSSIER

APRES CONSULTATION DE CONNAISSANCE ON OBTIENT:

	COMPREND	LE	DOSSIER
VS	COMPRENDRE(CONTENIR)	LE	DOSSIER

RESOLUTION DE LA RELATION SUIVANTE:
****

DANS	LES LIEUX	LOU#ES	
SP	LIEU	LOCATION|LOUANGE	

LES DIFFERENTS SENS DU PREMIER MOT SONT:

	LEUR FAMILLE EST:
LIEU	LIEU

LES DIFFERENTS SENS DU SECOND MOT SONT:

	LEUR FAMILLE EST:			
LOCATION	LOCATAIRE LOUABLE LOUEUSE	LOCATEUR LOUAGE	LOCATIF LOUER	LOCATION LOUEUR
LOUANGE	LOUABLE	LOUANGE	LOUER	

APRES CONSULTATION DE CONNAISSANCE ON OBTIENT:

DANS	LES LIEUX	LOU#ES	
SP	LIEU	LOCATION	

RESOLUTION DE LA RELATION SUIVANTE:
****

AU	R@EGLEMENTS	SANITAIRE	
SP	R#EGLEMENTER|R#EGLER(PAYER)	SANITAIRE	

LES DIFFERENTS SENS DU PREMIER MOT SONT:

	LEUR FAMILLE EST:			
R#EGLEMENTER	R#EGLEMENTAIRE R@EGLEMENT	R#EGLEMENTAIREMENT	R#EGLEMENTATION	R#EGLEMENTER
R#EGLER(PAYER)	R#EGLER	R@EGLEMENT		

LES DIFFERENTS SENS DU SECOND MOT SONT:

	LEUR FAMILLE EST:
SANITAIRE	SANITAIRE

APRES CONSULTATION DE CONNAISSANCE ON OBTIENT:

AU	R@EGLEMENTS	SANITAIRE	
SP	R#EGLEMENTER	SANITAIRE	

fig. 103.

CONCLUSION GENERALE

Nous allons aborder successivement trois aspects relatifs respectivement : aux **limites** des outils que nous avons construits, aux **applications** néanmoins possibles de notre travail, et enfin aux **espoirs** que l'on peut raisonnablement fonder sur notre approche pour aller plus loin dans l'investigation et l'analyse automatique de la langue naturelle.

I. Les limitations :

En commençant le présent travail nous avions en vue l'élaboration d'un algorithme permettant de construire toutes les connexions qui s'établissent entre les mots de la chaîne écrite. Parallèlement nous souhaitions que la représentation qui serait ainsi obtenue puisse permettre la détection de certaines synonymies lexico-syntaxiques. De surcroît enfin, nous voulions pouvoir appréhender toutes sortes de textes, sans contrainte aucune quant à leur nature ou à leur diversité tant dans le vocabulaire que dans les constructions syntaxiques.

Très vite l'on se rend compte, comme c'est souvent le cas, que de tels objectifs sont somme toute très ambitieux : les connexions sont tout d'abord fort nombreuses ; en outre leur détermination réclame que soient enregistrées un très grand nombre d'informations.

Ainsi qu'il est usuel d'opérer face aux problèmes complexes, nous avons alors émis des hypothèses et fait des choix simplificateurs. Ces simplifications devaient cependant être différentes de celles habituellement consenties. En particulier elles ne devaient pas porter sur la nature et la diversité des textes traités. Autrement dit le domaine appréhendé devait rester ouvert.

C'est donc sur la nature même de notre principal objectif, c'est à dire sur les connexions elles-mêmes que portèrent les limitations et approximations dont nous devions nous accommoder.

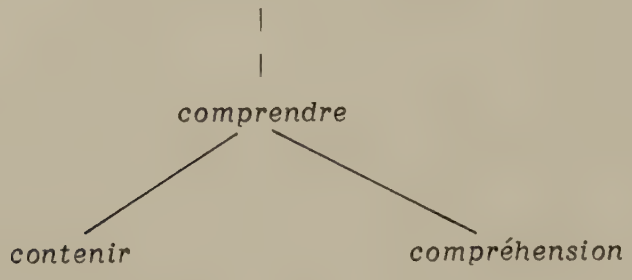
Elles furent doubles :

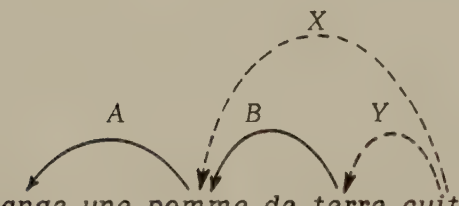
- En premier lieu sur **le nombre des connexions,** en ne cherchant à déterminer que les relations qui s'établissent entre les mots pleins. Par exemple,parmi les anaphoriques, seuls "qui" a été étudié. Ils sont pourtant à la fois nombreux et très répandus. Une autre limitation qui va dans le prolongement de la précédente, a été de délaisser, dans une première étape, les relations qui peuvent s'établir entre les éléments appartenant à des phrases différentes.

- En second lieu sur **la nature des connexions :** en ne retenant que des relations binaires d'une part, et en acceptant pour celles-ci une définition relativement large d'autre part.

• Insuffisance des RLS syntagmatiques binaires :

(a) *le conseil comprend le rapport*





(b) *l'artiste mange une pomme de terre cuite*

(J.B. Berthelin, 19)

Ces deux exemples montrent qu'il y a lieu, en (a), de faire intervenir une RLS ternaire pour résoudre la polysémie de "comprendre", et en (b) une RLS quaternaire pour éliminer la dépendance potentielle "Y".

Les deux exemples suivants :

(c) le conseil de classe comprend le professeur principal
(d) le lecteur de la thèse comprenant ces quatre phrases

où l'on a en (c) une ambiguïté lexicale et en (d) une ambiguïté lexico-syntaxique, montrent par contre qu'une telle solution peut elle-même s'avérer insuffisante.

Notons en outre que l'exemple (b) met en relief une autre faiblesse apparente du système, en relation avec la présente limitation : la non prise en compte des mots composés (mais voir plus loin, § III).

• Le deuxième point est relatif à la définition somme toute large des relations. Nous pensons notamment aux relations *Verbe - Complément Direct* et *Verbe - Complément Indirect* dont les frontières sont restées relativement perméables, ou encore, aux relations *Substantif - Substantif*, qui, finalement, peuvent inclure aussi bien des relations de "qualification" que des relations *Verbe - Sujet* ou *Verbe - Complément* (cf. ch. II, § V. 2. 3.) .

En revanche nous avons relevé le défi quant à l'acquisition automatique des informations nécessaires pour effectuer la reconnaissance des relations retenues et, partant, quant à l'analyse de larges données textuelles sans restriction sur leur nature ou diversité.

L'analyseur syntaxico-sémantique proposé se caractérise en effet par une acquisition automatique des informations linguistiques qui se retrouve à trois niveaux :

- sémantique : c'est alors l'apprentissage des RLS syntagmatiques ;
- morphologique : c'est alors la génération automatisée de familles de mots ;
- syntaxique enfin : c'est alors la simulation de règles transformationnelles au moyen d'un raisonnement par analogie rendant possible la détection de certaines paraphrases.

II. Les applications :

Cependant, bien que les limitations dont nous nous sommes accommodé soient de taille, nous pensons qu'il est néanmoins possible d'utiliser d'ores et déjà avec succès notre analyseur dans certaines applications telles que la documentation automatique ou encore la reconnaissance de la parole.

• ***Documentation automatique*** : pour cette application la représentation des données textuelles obtenues peut en effet se prêter aisément à une interrogation documentaire ; à condition toutefois d'y inclure certaines informations complémentaires, dont la construction est simple, telles que par exemple les numéros de documents, le rang des mots, leur appartenance aux documents, etc. D'ailleurs ce que nous avons déjà baptisé consultation de la connaissance dans le chapitre dernier n'était rien d'autre qu'une interrogation qui portait sur l'ensemble des RLS acquises à partir d'un lot de textes.

Le principe de l'interrogation documentaire est dès lors simple. Il consiste à retenir en outre le ou les numéro(s) de document(s) dont seraient issues les RLS qui ont permis l'analyse dépendancielle de la question ou qui y apparaissent simultanément. Nous dirions alors qu'un document (ou encore qu'un paragraphe ou même une phrase) répond d'autant mieux à une question que les occurrences (les mots mais aussi leurs relations) qui les composent respectivement sont proches.

Par exemple si dans une base de données textuelles nous avons des documents où figurent les phrases suivantes :

(1) *Le premier ministre est nommé par le président de la république.*

(2) La nomination du premier ministre par le président de la république a été ..

A la question :

Qui nomme le premier ministre ?

tout document qui contiendrait l'une ou l'autre des deux phrases précédentes serait alors sélectionné.

Dans notre représentation de ces phrases d'une part et de la question d'autre part, il y a en effet, à partir d'un certain niveau de profondeur paradigmatique, identité entre certaines des RLS qui les composent respectivement.

Une distance fondée sur le calcul des notes de validité des relations de dépendance de la question, mais qui pourrait également faire intervenir d'autres critères tels que les fréquences des mots ou des relations, pourrait permettre alors de fournir une réponse hiérarchisée.

C'est là le premier axe de recherche que nous allons explorer dans un avenir proche. Nous voudrions aussi offrir, outre une liste de documents dits réponses, une sorte de "résumé - réponse" qui serait constitué, dans une première étape, d'une juxtaposition de "phrases - réponses" issues des documents - réponses. L'exemple donné ci-dessus montre par ailleurs que nous pourrions également envisager, pour un certain type de questions, de fournir la réponse en clair ; ici en répondant "le président de la république".

• ***Reconnaissance automatique de la parole*** : Ici, au delà de la reconnaissance des phonèmes se pose le problème de la segmentation en mots de la chaîne phonétique qui a été reconnue. Nous supposerons pour ce qui nous concerne que cette reconnaissance phonétique a été, sinon totalement, du moins en grande partie, correctement effectuée. Notons d'ailleurs que là encore, ainsi qu'il est souvent de règle en Intelligence Artificielle, ce sont quelquefois les étapes considérées ultérieures du traitement (dont font partie nos algorithmes) qui peuvent aider à améliorer le bilan global de la reconnaissance. Toutefois il reste nécessaire qu'une reconnaissance phonétique "minimale" ait été au préablable effectuée. Il s'agit donc ensuite de la segmentation en mots de la chaîne phonétique supposée reconnue et, plus loin encore, de son "orthographication" automatique.

Ces deux problèmes très liés trouvent leur solution à tous les niveaux. Les RLS acquises peuvent précisément contribuer pour une large part à leur résolution au niveau sémantique.

Pour illustrer ceci nous allons considérer deux exemples tirés d'une expérience menée en collaboration avec le LIMSI (7, 8, 46, 85) .

Nous nous plaçons ici au terme de cette expérience, c'est à dire à l'issue des traitements lexico-syntaxiques. Le problème qui se pose revient alors en général à effectuer un choix lexical.

Exemple 1 : ləprefedəlasɛn

Pour cette chaîne phonétique l'heuristique de segmentation lexico-syntaxique propose trois orthographications potentielles :

(1) le préfet de la cène

(2) le préfet de la scène

(3) le préfet de la Seine

dont une seule est évidemment à retenir.

Cette ambiguïté rappelle, ainsi que nous pouvons le voir, celle que l'on rencontre en cas de polysémie (à un même signe, ici phonétique, correspondent plusieurs graphèmes).

Les mêmes méthodes de résolution peuvent dès lors être utilisées : la consultation de l'ensemble des RLS acquises peut en effet aider à faire un choix, dans le cas présent en accordant le plus fort poids à la troisième orthographication.

Exemple 2 : Il est relatif à la détermination de l'accord en genre et en nombre lors de l'orthographication.

ləperimɛtrdəprɔtɛksjɔ̃demɔnymãklase

A l'issue de l'anlyse lexico-syntaxique de cette chaîne, nous diposons des trois orthographications potentielles suivantes :

(1) le périmètre de protection des monuments classé

(2) le périmètre de protection des monuments classée

(3) le périmètre de protection des monuments classés

La résolution de cette ambiguïté passe par la détermination du mot auquel se rapporte l'adjectif postérieur "klase". Là encore les RLS couplées à leur mode d'utilisation (raisonnement par analogie) peuvent aider à lever l'ambiguïté.

• Nous avons cité deux applications, mais il est clair que ce choix n'est pas restrictif. Nous pensons en effet que notre outil est suffisamment général pour rendre un service à tout concepteur de système mettant en jeu le langage naturel : par exemple la traduction automatique (B. VAUQUOIS, 110(b)) , l'enseignement assisté par ordinateur (D. KAYSER, 74) , la recherche de stratègies d'inférence dans les sytèmes réels (consultation des pages professionnelles dans l'annuaire PTT) (J.B. BERTHELIN, 19) ; etc.

III. Les espoirs :

Nous allons reprendre deux points déjà évoqués dans l'introduction à ce travail et développés dans les chapitres III, IV et V : l'acquisition automatique d'une part et l'utilisation d'autre part des informations linguistiques.

a) Sur l'acquisition automatique de nouvelles informations linguistiques :

Nous pensons ici en particulier aux catégories sémantiques et aux mots composés (pris comme un tout) dont nous n'avons point tenu compte et qui pourtant sont à l'origine de certaines faiblesses du système.

Nous pouvons remarquer que pour établir ces informations, il faut au préalable avoir accumulé un certain nombre d'observations. C'est ainsi que M. GROSS note, d'une façon plus générale, que *"Toute construction théorique a toujours été précédée d'un long travail d'accumulation systématique de données"* (64, Introduction) (*).

Bien sûr la machine peut alors se révéler un auxiliaire précieux. Ceci à plus d'un titre :

- d'une part dans la phase d'accumulation ou de "catalogie" (L. MEHL, 87) amorcée ici par l'acquisition,
- d'autre part au cours de la structuration des données, c'est à dire dans ce repère lors de la détection des catégories ou des assemblages.

1°) Détection des catégories sémantiques :

A l'issue des traitements décrits dans le présent travail, nous disposons d'une connaissance sous forme de neuf matrices qui donnent respectivement pour chacune des neuf relations retenues des indications sur la possibilité qu'ont deux mots d'entretenir entre eux cette relation. En outre il est possible de passer d'une relation à l'autre par le biais des familles de mots (synonymies lexico-syntaxiques).

La figure 104 donne un extrait de ces matrices.

Ainsi que nous pouvons le voir, ces données constituent un matériau qui pourrait se prêter à une compilation (soit par le linguiste soit par la machine) qui laisserait se dégager des catégories sémantiques.

Par exemple l'ensemble des substantifs qui sont à la fois complément direct possible pour "assurer" et "amortir" pourraient recevoir l'étiquette "BIEN MATERIEL".

(*) Il ne faut cependant pas en déduire que certaines théories resteront impossibles à élaborer par suite d'un trop grand nombre de faits qui devraient être observés. Voir la discussion de ce problème dans *"Science et Méthode"* H. POINCARE, ch. I.

209	VS	2938	ACCORDER		APR-ETNE
210	VS	3146	ACCORDER		ARR-ETNE
211	VS	5310	ACCORDER		MINISIRE
212	VS	7224	ACCORDER		MINISTRE
790	VS	9666	COMPRENDRE		ACTE
791	VS	2055	COMPRENDRE		DOSSIER
792	VS	2490	COMPRENDRE		DOSSIER
793	VS	8823	COMPRENDRE		ENSEMBLE
794	VS	8779	COMPRENDRE		ENSEMBLE
795	VS	1740	COMPRENDRE		SECTION
796	VS	1738	COMPRENDRE		SECTION
797	VS	4131	CONFRONETTRE	DE	GAZ
889	VC	1526	CONSENTIR	AUX	CONECHANGISTE
890	VC	1527	CONSENTIR	AUX	CONECHANGISTE
891	VC	5575	CONSENTIR	AUX	PARTICULIER
892	VC	12858	CONSENTIR	DANS	CONDITION
893	VC	8512	CONSENTIR	POUR	DURNEE
894	VC	8397	CONSENTIR	POUR	DURREE
895	VC	7799	CONSENTIR	POUR	DURREE
1170	VC	8049	DESTINER	ƏA	ACCESSION
1171	VC	6493	DESTINER	ƏA	HABITATION
1172	VC	7936	DESTINER	ƏA	LOCATION
1173	VC	8002	DESTINER	ƏA	LOCATION
1174	VC	4794	DESTINER	ƏA	RECEVOIR
1761	SA	1532	LOGEMENT		ANCIEN
1762	SA	1533	LOGEMENT		ANCIEN
1763	SA	1479	LOGEMENT		PETIT
1764	SA	1478	LOGEMENT		PETIT
1765	SP	3610	LOGEMENT		₩ELOIGNER
1766	SP	1945	LOGEMENT		₩EQUIPER
1767	SP	1949	LOGEMENT		₩EQUIPER
1768	SP	1456	LOGEMENT		ACHEVER
1769	SP	1448	LOGEMENT		ACHEVER
1770	SP	1457	LOGEMENT		ACHEVER
1771	SP	1449	LOGEMENT		ACHEVER
1772	SP	6929	LOGEMENT		CONSTRUIRE
1773	SP	1423	LOGEMENT		CONSTRUIRE
1774	SP	1424	LOGEMENT		CONSTRUIRE
1775	SP	8854	LOGEMENT		₩₩EPASSER
1776	SP	8879	LOGEMENT		₩₩EPASSER
1777	SP	7947	LOGEMENT		DESTINER
1778	SP	8048	LOGEMENT		DESTINER
1779	SP	7935	LOGEMENT		DESTINER
1780	SP	1922	LOGEMENT		HABITABLE
1781	SP	3610	LOGEMENT		HABITER
1782	SP	1346	LOGEMENT		SITUER
1783	SP	1347	LOGEMENT		SITUER

440	SS	5637	ALINNEA	DE	ARTICLE
441	SS	8113	ALINNEA	DE	ARTICLE
442	SS	6849	APPLICATION	DE	ARTICLE
443	SS	9391	APPLICATION	DE	ARTICLE
444	SS	12723	COMMENTAIRE	DE	ARTICLE
445	SS	6561	DISPOSITION	DE	ARTICLE
446	SS	7216	DISPOSITION	DE	ARTICLE
447	SS	6194	PARAGRAPHE	DE	ARTICLE
448	SS	6577	PARAGRAPHE	DE	ARTICLE
449	SS	6592	PARAGRAPHE	DE	ARTICLE
450	SS	7652	PARAGRAPHE	DE	ARTICLE
451	SS	9578	COMMENTAIRE	DES	ARTICLE
452	SS	8568	PARAGRAPHE	DU	ARTICLE
453	VC	10198	DNEFINIR	ƏA	ARTICLE
454	VC	1338	DNEFINIR	ƏA	ARTICLE
455	VC	1337	DNEFINIR	ƏA	ARTICLE
456	VC	8450	FIXER	ƏA	ARTICLE
457	VC	9220	FIXER	ƏA	ARTICLE
458	VC	8642	FIXER	ƏA	ARTICLE
459	VC	8304	FIXER	ƏA	ARTICLE
460	VC	1293	PRNEVOIR	ƏA	ARTICLE
461	VC	1294	PRNEVOIR	ƏA	ARTICLE
462	VC	6637	PRNEVOIR	ƏA	ARTICLE
463	VC	9018	PRNEVOIR	ƏA	ARTICLE
464	VC	9047	PRNEVOIR	ƏA	ARTICLE
465	VC	9237	PRNEVOIR	ƏA	ARTICLE
466	VC	10545	PRNEVOIR	ƏA	ARTICLE
467	VC	10757	PRNEVOIR	ƏA	ARTICLE
468	VC	8624	PRNEVOIR	ƏA	ARTICLE
469	VC	1898	VISER	ƏA	ARTICLE
470	VC	1902	VISER	ƏA	ARTICLE
471	VC	7969	VISER	ƏA	ARTICLE
472	VC	2826	VISER	ƏA	ARTICLE
473	VC	4450	VISER	ƏA	ARTICLE
474	VC	8891	VISER	ƏA	ARTICLE
475	VC	9651	₩ENUMERER	AUX	ARTICLE
476	VC	12862	DNEFINIR	AUX	ARTICLE
477	VC	5293	PRNEVOIR	AUX	ARTICLE
478	VC	12951	PRNEVOIR	AUX	ARTICLE
479	VC	6674	VISER	AUX	ARTICLE
1980	SS	1704	CONSEIL	DU	MINISTƏERE
1981	SS	1702	CONSEIL	DU	MINISTƏERE
1982	SS	1604	CONSEIL	DU	MINISTƏERE
1983	SS	1637	CONSEIL	DU	MINISTƏERE
1984	SS	1603	CONSEIL	DU	MINISTƏERE
1985	SS	1883	AUTORISATION	DES	MINISIRE
1986	SS	1879	AUTORISATION	DES	MINISIRE
1987	SS	3527	ACCORD	DU	MINISIRE
1988	SS	1647	ARR-ETNE	DU	MINISIRE
1989	SS	3492	ARR-ETNE	DU	MINISIRE
1990	SS	1649	ARR-ETNE	DU	MINISIRE
1991	SS	4363	CIRCULAIRE	DU	MINISIRE
1992	VC	6249	TRANSMETTRE	AU	MINISTRE
1993	VC	7586	TRANSMETTRE	AU	MINISTRE
1994	VS	7224	ACCORDER		MINISTRE
1995	VS	5310	ACCORDER		MINISIRE
1996	VS	1705	APPELER		MINISIRE
1997	VS	1701	APPELER		MINISIRE
1998	VS	7449	CHOISIR		MINISIRE
1999	VS	2745	DNELIVRER		MINISIRE
2000	VS	7410	DNESIGNER		MINISIRE
2001	VS	3323	PRENDRE		MINISIRE

fig. 104. Extrait de quelques lignes
ou colonnes des matrices
Verbe - Sujet (**VS**), Verbe - Complément Indirect (notée **VC**),
Verbe - Complément direct (notée **SP**), Substantif - Substantif (**SS**),
Substantif - Adjectif - Antérieur (**SA**).

Les procédés qui permettraient de déterminer automatiquement de telles catégories n'ont pas encore fait l'objet de notre part d'une étude concrète. Mais nous pouvons d'ores et déjà conjecturer les deux points suivants :

- Les catégories que nous obtiendrons automatiquement ne correspondront pas toujours forcément aux catégories qui nous sont familières.
  Ceci est d'une part prévisible étant donné que les systèmes automatiques obtiennent fréquemment les mêmes résultats que les humains en passant par des intermédiaires différents ; d'autre part on peut envisager que les catégories ainsi obtenues pourront présenter pour le linguiste ou le psychologue un certain intérêt.
  Un autre avantage de cette automatisation est que l'on pourra recalculer, en fonction d'un changement de situation, tout ou partie de l'ensemble des catégories.
- En second lieu, il convient de souligner que l'acquisition de catégories sémantiques ne remet pas en question la nécessité des RLS. En effet, ces deux informations correspondent à des besoins différents.
  Par exemple si nous considérons la phrase "je vous désire une bonne année" (J. PITRAT, 95, p. 881) , nous avons besoin d'une part de comprendre ce que veut dire cette phrase, d'autre part de décider si elle est acceptable en français.
  Détaillons sur un schéma les deux types de connaissance que nous ferons intervenir pour cet exemple :

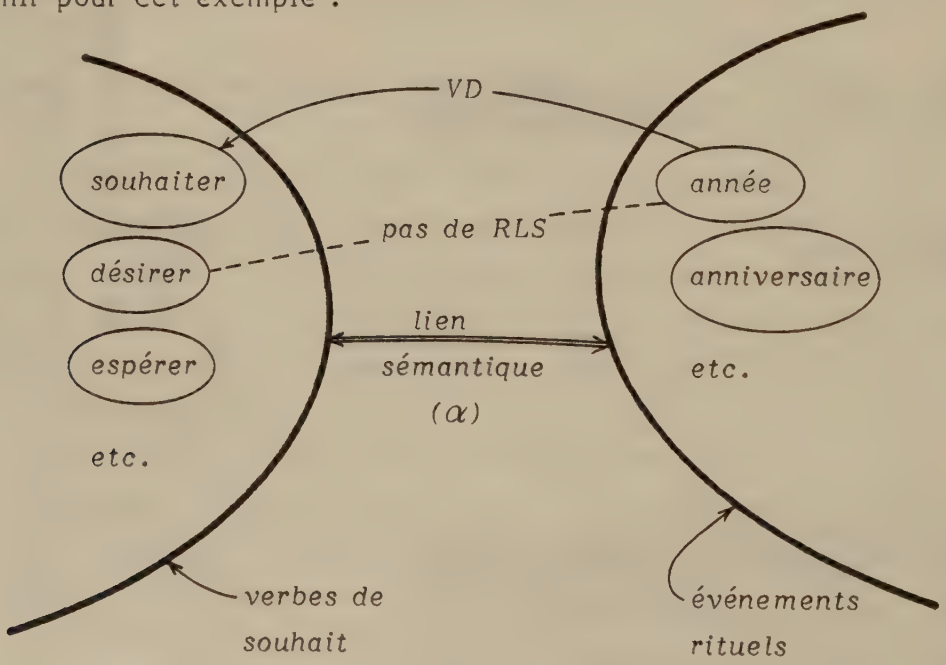


fig. 105.

Ici le lien sémantique α permet de comprendre que la phrase exprime des voeux à l'occasion d'un événement rituel, et l'absence de la RLS *(désirer, année, VD)*

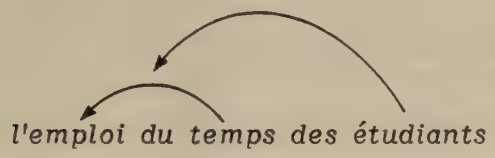
suggère que la phrase n'est pas possible en français. Nous rejoignons ainsi A.K. ZHOLKOVSKIJ et I.A. MEL'CHUK qui notent que *"ce n'est qu'en disposant d'une large information de ce genre* (i.e. relations lexico-sémantiques) *que l'on peut avoir une périphrase riche et une expression idiomatique du sens en question"*. (115, p. 14) .

2°) Détection des mots composés :

Là aussi, à partir du même matériau il est possible de détecter ce qu'il est convenu d'appeler mots composés.

On peut par exemple envisager, en se fondant sur la fréquence d'une RLS et sur des critères de contiguïté, de faire ressortir des combinaisons stables. La figure 106 offre quelques exemples.

C'est munis de cette information qu'il sera alors possible de retenir pour la phrase :



le fléchage indiqué.

b) Sur l'utilisation de cette nouvelle connaissance :

1°) des catégories :

- Le pouvoir inférentiel des RLS peut être accrû. En effet jusqu'à présent le contrôle sémantique s'est opéré à l'aide de RLS ; mais les catégories sémantiques étant plus larges que les familles de mots existantes, la représentation de la connaissance aura plus de profondeur en terme paradigmatique ; cela pourra conduire donc à éliminer certains silences.

  Par exemple, la phrase :

  Le commandant du navire qui écrit ...

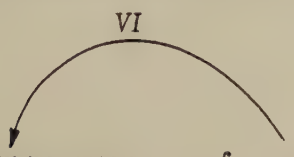
pourrait être correctement fléchée en utilisant d'une part la RLS (*capitaine, écrire, VS*) et d'autre part une catégorie sémantique qui inclurait "capitaine" et "commandant".

1893	SS	1613	MINISTƏERE	DE	CONSTRUCTION
1894	SS	7541	MINISTƏERE	DE	CONSTRUCTION
1895	SS	1607	MINISTƏERE	DE	CONSTRUCTION
1896	SS	1640	MINISTƏERE	DE	CONSTRUCTION
1897	SS	1612	MINISTƏERE	DE	CONSTRUCTION
1898	SS	1707	MINISTƏERE	DE	CONSTRUCTION
1901	SS	7588	MINISTRE	DE	CONSTRUCTION
1902	SS	5313	MINISTRE	DE	CONSTRUCTION
1903	SS	6251	MINISTRE	DE	CONSTRUCTION
1904	SS	6619	MINISTRE	DE	CONSTRUCTION
1905	SS	7227	MINISTRE	DE	CONSTRUCTION
1906	SS	7405	MINISTRE	DE	CONSTRUCTION
1907	SS	7452	MINISTRE	DE	CONSTRUCTION
1908	SS	7308	MINISTRE	DE	CONSTRUCTION
1909	SS	4365	MINISTRE	DE	CONSTRUCTION
1910	SS	1651	MINISTRE	DE	CONSTRUCTION
1911	SS	6257	MINISTRE	DE	INDUSTRIE
1912	SS	3326	MINISTRE	DE	RECONSTRUCTION
1913	SS	3494	MINISTRE	DE	RECONSTRUCTION
1914	SS	3535	MINISTRE	DE	SANT#E
1915	SS	3468	MINISTRE	DE	SANT#E
1916	SS	4371	MINISTRE	DE	SANT#E
1917	SS	3171	MINORATION	DE	DENSIT#E
2193	SS	3521	PERMIS	DE	CONSTRUIRE
2194	SS	3148	PERMIS	DE	CONSTRUIRE
2195	SS	3098	PERMIS	DE	CONSTRUIRE
2196	SS	2958	PERMIS	DE	CONSTRUIRE
2197	SS	3203	PERMIS	DE	CONSTRUIRE
2198	SS	1593	PERMIS	DE	CONSTRUIRE
2199	SS	6801	PERMIS	DE	CONSTRUIRE
2200	SS	6139	PERMIS	DE	CONSTRUIRE
2201	SS	6865	PERMIS	DE	CONSTRUIRE
2202	SS	6888	PERMIS	DE	CONSTRUIRE
2203	SS	5857	PERMIS	DE	CONSTRUIRE
2204	SS	1592	PERMIS	DE	CONSTRUIRE
2205	SS	3488	PERMIS	DE	CONSTRUIRE
2206	SS	6830	PERMIS	DE	CONSTRUIRE
2207	SS	3243	PERMIS	DE	CONSTRUIRE
2208	SS	6078	PERMIS	DE	CONSTRUIRE
2209	SS	6549	PERMIS	DE	CONSTRUIRE
2210	SS	2052	PERMIS	DE	CONSTRUIRE
2211	SS	6658	PERMIS	DE	CONSTRUIRE
2212	SS	3129	PERMIS	DE	CONSTRUIRE
2213	SS	6709	PERMIS	DE	CONSTRUIRE
2214	SS	7238	PERMIS	DE	CONSTRUIRE
2818	SP	10905	SOCI#ET#E		ANONYME
2819	SP	10873	SOCI#ET#E		ANONYME
2820	SP	10817	SOCI#ET#E		ANONYME
2821	SP	1400	SOCI#ET#E		ANONYME
2822	SP	10791	SOCI#ET#E		ANONYME
2823	SP	10702	SOCI#ET#E		ANONYME
2824	SP	4407	SOCI#ET#E		ANONYME
2825	SP	5923	SOCI#ET#E		ANONYME
2826	SP	10224	SOCI#ET#E		ANONYME
2827	SP	10623	SOCI#ET#E		ANONYME
2828	SP	11120	SOCI#ET#E		ANONYME
2829	SP	10379	SOCI#ET#E		ANONYME
2830	SP	10776	SOCI#ET#E		ANONYME
2831	SP	5972	SOCI#ET#E		ANONYME
2832	SP	10159	SOCI#ET#E		ANONYME
2833	SP	10459	SOCI#ET#E		ANONYME
2834	SP	10533	SOCI#ET#E		ANONYME
2835	SP	4497	SOCI#ET#E		ANONYME
2836	SP	5934	SOCI#ET#E		CIVIL
2837	SP	5910	SOCI#ET#E		CIVIL
2838	SP	6043	SOCI#ET#E		CIVIL
2839	SP	5999	SOCI#ET#E		CIVIL
2840	SP	6037	SOCI#ET#E		CIVIL
2979	SA	7737	TERME		LONG
2980	SA	7899	TERME		LONG
2981	SA	7918	TERME		LONG
2982	SA	8481	TERME		LONG
2983	SA	7768	TERME		LONG
2984	SA	7673	TERME		LONG
2985	SA	7862	TERME		LONG
2986	SA	8354	TERME		LONG
2987	SA	8214	TERME		LONG
2988	SA	7792	TERME		MOYEN
2989	SA	8508	TERME		MOYEN
2990	SA	8339	TERME		MOYEN
2991	SA	7912	TERME		MOYEN
2992	SA	8437	TERME		MOYEN
2993	SA	8141	TERME		MOYEN
2994	SA	7716	TERME		MOYEN

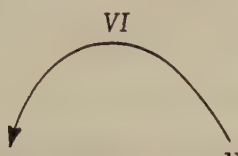
fig. 106. Extrait de quelques lignes ou colonnes des matrices Substantif - Substantif (**SS**), substantif - Adjectif Postérieur (**SP**) et Substantif - Adjectif Antérieur (**SA**). On voit clairement apparaître des mots composés.

- Il serait également possible d'affiner les RLS syntagmatiques, ceci dans le but de compenser l'ambiguïté prépositionnelle et d'autres du même type. Cela permettrait alors de faire un pas en direction d'une grammaire de cas.

Considérons, par exemple, les phrases :



(1) Il mange avec une fourchette



(2) Il mange avec un collègue

Les deux relations mettant en jeu un complément indirect introduit par "avec" sont pour le moment confondues. En utilisant le fait que "fourchette" a la catégorie "OBJET" et "collègue" la catégorie "HUMAIN" il serait donc possible de les distinguer automatiquement.

Des situations analogues peuvent se présenter dans l'interprétation du sujet. Par exemple le sujet de guérir peut correspondre à plusieurs cas sémantiques différents, comme le montrent les quatre exemples suivants :

(a) Le malade guérit
(b) Le rhume guérit
(c) Le médecin guérit le malade
(d) L'aspirine guérit le rhume

(J.B. BERTHELIN, 19)

Là aussi la connaissance des catégories sémantiques nous sortirait d'affaire.

Ceci peut présenter un intérêt dans une application documentaire où le but serait de répondre en clair. Par exemple à la question *"Qui guérit le rhume ?"* il s'agit évidemment d'éviter (a), (b) et (d) qui constitueraient des bruits. Inversement, si l'on demande *"Avec quoi guérit-on le rhume ?"* il faut éviter le silence sur (d).

2°) des mots composés :

Restons dans la perspective documentaire pour illustrer notre propos.

Si l'on demande l'adresse d'un hôtel, il est évidemment préférable que le système ne fasse pas figurer l'adresse de l'hôtel de ville parmi les réponses.

Signalons toutefois que le problème est loin d'être résolu car il comporte un aspect difficile : comment peut-on déterminer automatiquement si le mot composé "X de Y" est ou n'est pas spécifique de X ?

IV. Caractère évolutif du système :

Tout ce que nous venons de développer laisse transparaître un aspect essentiel du système, qui est d'élaborer toujours un type de connaissance à partir d'un autre, plus directement accessible par l'expérience. C'est ce qui confère à notre système une capacité évolutive, c'est à dire le moyen de se comporter plus finement au fur et à mesure qu'il a traité plus de données. Cette évolution n'est cependant pas entièrement automatique, mais dépend de la fourniture de nouveaux algorithmes à chaque nouvelle configuration de l'état des connaissances. Autrement dit le "mode d'emploi" des nouvelles connaissances n'est pas déterminé automatiquement, quoi que celles que nous avons décrites soient, par exemple, en partie destinées à prolonger l'axe paradigmatique. De manière plus générale, un accroissement des connaissances ou un affinement des résultats ne demande qu'un investissement local au niveau des algorithmes, et non un remaniement global du système.

Il n'en reste pas moins que nous avons dû accepter consciemment, en créant ce système, le fait que certaines difficultés lui survivraient, même à long terme : ne serait-ce que l'usage des connaissances pragmatiques, que nous n'avons même pas abordé ; en outre, il est clair que toutes les connaissances nécessaires à l'analyse du langage ne peuvent être apprises à partir des seuls textes. Enfin, l'analyse d'un texte suppose de connecter entre elles les phrases qui le composent, point que, rapellons-le, nous n'avons pas non plus étudié.

On pourrait trouver, bien sûr, nombre d'exemples analogues, auxquels se sont attaquées de nombreuses recherches. Cependant nous pensons avoir réalisé un système relativement compatible avec ceux qui existent : en effet, ainsi que nous l'avons souligné au chapitre I, notre travail, nous semble-t-il, se situe en amont des recherches généralement menées dans le domaine de l'Intelligence Artificielle.

Par ailleurs, si nous passons à une perspective plus large, il nous semble que, d'une part, ce que nous avons appelé RLS pourrait constituer un matériau, obtenu par une démarche empirique, et ayant des chances de se révéler utile à l'élaboration de l'édifice théorique de la sémantique ; et que d'autre part, notre système pourrait s'intégrer à l'entreprise très vaste que représenterait

la nécessaire coopération entre les diverses disciplines qui tendent à apporter des éléments de réponse au problème de la compréhension du langage naturel.

Annexes

LEXIQUE DES FIGURES

Figures		Pages
1.	Graphe donnant les RLS possibles	37
2.	Représentation graphique d'une RLS	39
3.	Matrice des synonymies lexico-syntaxiques potentielles.	54
4.	Graphe des synonymies lexico-syntaxiques potentielles.	55
5.	Non-transitivité des équivalences potentielles.	55
6.	Forme sous laquelle se présente le texte à l'entrée du système présenté.	59
7.	Agencement des traitements morpho-grammaticaux et traitements syntaxico-sémantiques, objet du présent travail.	60
8.	Représentation matricielle de l'agencement des filtres élémentaires pour la reconnaissance des chaînes nominales.	76
9.	Représentation à l'aide d'un réseau de transition de ces mêmes agencements.	76
10.	Matrice de reconnaissance des chaînes nominales.	78
11.	Réseau de transition correspondant à cette matrice.	79
12.	Matrice de reconnaissance de chaînes nominales ayant servie pour l'expérimentation.	82
13.	Exemples de reconnaissance de chaînes nominales sur un texte scientifique.	83
14.	Exemples de reconnaissance de chaînes nominales sur un texte juridique.	84
15.	Performances de l'algorithme de reconnaissances des chaînes nominales.	81
16.	Matrice de reconnaissance des chaînes verbales.	86
17.	Exemples de segmentation de phrase : reconnaissance de chaînes nominales et verbales.	88 & 89
18.	Performances de l'algorithme de reconnaissance des chaînes verbales.	87
19.	Un exemple de non-reconnaissance de chaînes verbales.	90
20.	Organigramme des étapes de segmentation.	91
21.	Exemple de représentation dépendancielle souhaitée.	93
22.	Représentations dépendancielles potentielles associées à la phrase :"Vente de produits de beauté aux algues marines".	94
23.	Les quatre situations possibles selon qu'il y a présence ou non des bruit et silence.	95

Figures — Pages

24. Typologie des diverses relations satellites. 98
25. Les différents substantifs selon qu'ils sont ou non déterminés, précédés ou non d'une préposition. 102
26. Exemples de reconnaissance de relations satellites. 104
27. Reconnaissance des relations Subtantif - adjectif Antéposé (SA). 106
28. Performances de la procédure de reconnaissance des relations (SA). 107
29. Reconnaissance des relations Substantif - adjectif Postposé (SP). 110
30. Performances de la procédure de reconnaissance des relations (SP). 109
31. Silences enregistrés lors de la reconnaissance des relations (SS_p). 112
32. Silences enregistrés lors de la reconnaissance des relations (SS_p). 112
33. Performances de la procédure de reconnaissance des relations (SS_p). 113
34. Reconnaissance des relations Substantif - Substantif (SS_p). 114
35. Reconnaissance des relations Verbe - Sujet (VS) et Verbe - complément Direct (VD). 117
36. Reconnaissance des relations Verbe - complément Direct (VD) et Verbe - complément Indirect (VI_p). 118
37. Performances des procédures de reconnaissance des relations VS, VD et VI_p. 119
38. Reconnaissance des relations Verbe - Adverbe (VA). 120
39. Silences volontaires dûs à la non-reconnaissance des relations Adjectif - Complément d'adjectif (AC_p). 120
40. Construction des relations de dépendance potentielles internes aux chaînes nominales. 122
41. Représentation dépendancielle de la phrase : "Le professeur fait réciter la leçon à l'élève". 125
42. Représentation dépendancielle de la phrase : "L'organisateur de la conférence voit arriver les participants". 126
43. Reconnaissance de l'élément central et du type de la chaîne verbale. 130
44. Performances de l'analyse des chaînes verbales. 129
45. Catégorisation des chaînes verbales. 134
46. Reconnaissance du sujet dans le cas des formes à l'indicatif, au subjonctif ou au conditionnel (formes MP). 136 & 137
47. Performances de la procédure de reconnaissance du sujet dans les formes MP. 135
48. Bruits et silences enregistrés lors de la reconnaissance du sujet dans les formes MP. 139
49. Performances de la procédure de reconnaissance du sujet dans les formes MIP. 140

Figures

		Pages
50.	Reconnaissance des sujets potentiels dans les formes MIP.	141
51.	Silences enregistrés lors de la reconnaissance du sujet dans les formes MIP.	142
52.	Silences provoqués par la non-recherche des sujets potentiels pour les formes MII.	144
53.	Performances de la procédure de reconnaissance du sujet dans les formes passives.	145
54.	Reconnaissance des relations VS dans les formes passives.	146
55.	Bruits totaux enregistrés lors de la reconnaissance des relations VS dans les formes passives.	147
56.	Performances hypothétiques de la procédure de reconnaissance des relations VS si "de" avait été retenu comme introducteur du sujet dans les formes passives.	145
57.	Performances de la procédure de reconnaissance des relations VD dans les formes actives.	148
58.	Reconnaissance des relations VD dans les formes actives.	149
59.	Silences totaux enregistrés lors de la reconnaissance des relations VD dans les formes actives.	150
60.	Performances de la procédure de reconnaissance des relations VD dans les formes passives MP.	151
61.	Reconnaissance des relations VD dan les formes passives.	152
62.	Performances de la procédure de reconnaissance des relations VI_p dans les formes passives.	153
63.	Silence total enregistré lors de la reconnaissance des relations VI_p dans les formes passives.	154
64.	Reconnaissance des relations VI_p dans les formes passives.	155
65.	Performances de la procédure de reconnaissance des antécédents potentiels de l'anaphorique "qui".	157
66.	Reconnaissance des liens anaphoriques.	158 & 159
67.	Silences totaux enregistrés lors de la reconnaissance des antécédents de l'anaphorique "qui"	160
68.	Enchaînement des procédures pour l'obtention des graphes de dépendance potentiels d'une phrase.	163
69.	Analyse de certaines phrases extraites des textes sur la règlementation de l'urbanisme.	164 à 172
70.	Extrait du dictionnaire ayant servi à l'expérience.	180 et 181
71.	Ensemble des préfixes utilisés dans l'expérience.	182
72.	Ensembles des suffixes utilisés dans l'expérience.	183
73.	Décomposition multiple des mots en radical + suffixe.	185
74.	Extrait des familles primitives suffixales.	187

Figures — Pages

75. Quelques schémas dérivationnels. 190
76. Matrice de corrélation binaire des suffixes. 194
77. Matrice de compatibilité des suffixes. 195
78. Matrice d'incompatibilité des suffixes. 196
79. Extrait des familles primitives scindées. 204
80. Extrait des familles de mots obtenues après réunion transitive. 210
81. Ensemble des suffixes et suffixes composés. 211
82. Quelques familles suffixales obtenues automatiquement. 213
83. Quelques familles primitives préfixales. 214
84. Exemples de gestion des familles de mots. 216
85. Organigramme général et résultats chiffrés de l'algorithme de reconnaissance des familles de mots. 217, 218 et 222
86. Description paradigmatique de quelques phrases. 223
87. Taux des relations de dépendance ayant été décrétées Valides ou Douteuses à l'issue des traitements syntaxiques. 231
88. Performances du traitement syntaxique. 232
89. Performances détaillées du traitement syntaxique. 234
90. Représentation d'une relation de dépendance. 233
91. Représentation de "affichages muraux". 235
92. Représentation de la phrase "Les affichages muraux sont interdits". 235
93. Schéma du raisonnement par analogie. 236
94. Exemple pour illustrer le mécanisme de résolution des ambiguïtés syntaxiques. 238
95. Détail de la figure 94. 239
96. Table de notation. 241
97. Organigramme général du traitement syntaxico-sémantique. 243
98. Quelques exemples de résolution d'ambiguïtés syntaxiques faisant intervenir la connaissance. 245 à 247
99. Performances du traitement sémantique. 248
100. Performances globales du traitement syntaxico-sémantique. 249
101. Performances de l'algorithme d'analyse syntaxico-sémantique au cours du temps. 251
102. Un exemple détaillé de résolution d'ambiguïté lexicale (sans ajout dans la connaissance). 256
103. Exemples de résolution d'ambiguïtés lexicales (avec ajouts volontaires dans la connaissance). 257

Figures

Pages

104. Extrait de quelques lignes ou colonnes des matrices VS, VI_p, VD, SS_p et SA. 265

105. Schéma montrant sur un exemple que catégories sémantiques et RLS sont utiles. 266

106. Extrait de quelques lignes ou colonnes des matrices SS_p, SP et SA. 268

LEXIQUE TERMINOLOGIQUE

Bruit - Silence : A la suite de l'informatique documentaire, nous empruntons à la théorie du signal les deux notions fort utiles de "bruit" et de "silence". On parle de bruit quand quelque chose est capté alors que rien n'est émis, et de silence quand il y a émission et non réception.

Chaîne nominale - chaîne verbale :

Le vocable "syntagme - verbal" inclut généralement, de façon facultative ou obligatoire selon la nature du verbe (verbe ou copule), le syntagme adjectival, nominal et/ou prépositionnel qui suit le verbe (J. DUBOIS, 51 & 52) .

Dans ce travail nous introduisons le vocable "chaîne verbale" (CV) pour désigner exclusivement le syntagme verbal dépouillé de ses expansions nominales.

A l'opposé, étant souvent conduit à confondre, sur le plan du traitement, syntagme nominal et syntagme prépositionnel, nous introduirons alors le vocable "chaîne nominale" (CN) pour désigner indifféremment l'un ou l'autre de ces syntagmes.

Famille de mots : *"est l'ensemble des mots qui peuvent se grouper autour d'un radical commun d'où ils ont été tirés par la dérivation ou par la composition"*. (M. GREVISSE, 62, p. 105) .

Filtre : Opérateur qui permet d'extraire, d'un texte, toutes les occurrences d'une configuration de mots définie par des indications morphologiques, syntaxiques ou mixtes.

Mot plein - Mot vide :

Nous reprenons telles quelles (au sens de Tesnière) ces notions devenues d'usage courant ; rappelons que nous désignons par "mots pleins" les substantifs, verbes, adjectifs et adverbes, et par "mots vides" ceux qui n'entrent dans aucune de ces quatre catégories.

Relation lexicale sémantique (RLS) :

- **syntagmatique :** Relation définie sur un n-uplet de mots pleins extraits d'un texte et vérifiant certaines conditions morpho-syntaxiques.
- **paradigmatique :** Relation qui s'établit entre les unités lexicales qui sont susceptibles de commuter.

Schéma dérivationnel :

C'est un ensemble de suffixes qui peuvent être rapportés à un même radical.

CODAGE DES LETTRES ACCENTUEES

Les logiciels que nous avons construits ont été développés et expérimentés, en partie, sur les matériels du CIERCE (CNRS), et, dans une plus large part, sur les matériels de la CISI à Saclay, où ils sont actuellement implémentés.

Au moment où nous avons effectué nos expériences, les matériels de saisie et d'édition dont nous disposions n'étaient pas pourvus des lettres accentuées, aussi avons-nous adopté les conventions de codage suivantes :

accent aigu	:	≠	: ≠E = é
accent grave	:	@	: @A = à, @E = è, ...
accent circonflexe	:	¬	: ¬E = ê, ¬I = î, ...
tréma	:	<	: <I = ï, ...
cédille	:	$	: $C = ç
majuscule	:	&	: Paris = &PARIS

Ajoutons que le langage de programmation que nous avons utilisé est le PL/1.

BIBLIOGRAPHIE

1 - ABELSON R.
The structure of belief systems in SCHANK R.C., COLBY K.M. (eds) "computer models of thought and langage" Freeman - San Francisco - 1973 - pp. 287-340

2 - A. ANDREEWSKY, C.FLUHR
Apprentissage, Analyse automatique du langage, application à la documentation.
Dunod - Documents de linguistique quantitative n° 21 (1973).

3 - A. ANDREESWSKY, C. FLUHR
A learning method for natural language processing and application to information retrieval.
IFIP congress, août 1974.

4 - A. ANDREEWSKY, F. COMBRISSON, C. FLUHR
Le problème de l'identification automatique des concepts.
Note C.E.A - N-1816, sept. 1975.

5 - A. ANDREEWSKY, F. DEBILI, C. FLUHR, Y. HLAL, L. IVANOVA
Algorithme d'apprentissage pour la construction d'une syntaxe permettant l'analyse grammaticale de la langue russe
Note C.E.A - N-1922, Décembre 1976, pp. 132.

6(a)- A. ANDREEWSKY, F. DEBILI, C.FLUHR
Problems of computational learning of semantic relations
International seminar on intelligent question-answering and data base systems, Bonas (Gers) (juin 1977, pp. 9).

6(b)- A. ANDREEWSKY, F. DEBILI, C. FLUHR
Computational learning of semantic lexical relations for the generation and automatic analysis of content (pp. 667-673)
IFIP, congres Toronto, août 1977.

7 - A. ANDREEWSKY, F. DEBILI, C.FLUHR, J.S. LIENARD, J.J. MARIANI
Une expérience d'aide linguistique à la reconnaissance automatique de la parole
Note C.E.A - N-2055, Octobre 1978, pp. 10

8 - A. ANDREEWSKY, J.P. BINQUET, F. DEBILI, Y. HLAL, J.S. LIENARD, J.J. MARIA
Les dictionnaires en formes complètes et leur utilisation dans la transformation lexicale et syntaxique correcte de chaînes phonétiques.
Journées d'étude sur la parole - mai-juin 1979 - publié par le Groupe de Communication parlée du GALF, pp. 10.

9 - A. ANDREEWSKY, F. DEBILI, C. FLUHR
Apprentissage - Syntaxe - Sémantique lexicale
Revue du Palais de la Découverte - Vol. 9 - N°83

10 - A. ANDREEWSKY, J.P. BINQUET, F. DEBILI, C. FLUHR, B. POUDEROUX
Le traitement linguistique et statistique des textes et son application dans le domaine de la documentation juridique
Symposium sur l'informatique juridique, Juillet 1981, Salonique.

11 - A. ANDREEWSKY, F. DEBILI, C. FLUHR, B. POUDEROUX
SPIRIT, a Syntactical and Probabilistic System for Indexing and Retrieval of Information from texts.
EUSIDIC Conference 81, Bern, Switzerland, 13-15 october 1981.

12 - Ju.D. APRESJAN
Eléments sur les idées et les méthodes de la linguistique structurale contemporaine.
Monographies de linguistique mathématique n°5, Dunod 1973.

13 - J.C. BASSANO
Acquisition de connaissances à partir de textes : une expérimentation en documentation automatique?
LRI - ORSAY Rapport de recherche N°41 - Mai 1979.

14 - J.C. BASSANO
L'analogie dans un algorithme de documentation automatique.
Journées de travail sur l'analogie à la MSH PARIS, 21 et 22 Février 1980.

15 - J.C. BASSANO et A. HERZALLAH
Composante linguistique pour un système de documentation automatique.
Congrès AFCET, NANCY, Sept. 1981.

16 - N. BELY, A. BORILLO, J. VIRBEL, N. SIOT-DECAUVILLE
Procédures d'analyse sémantique appliquées à la documentation scientifique.
Gauthiers-Villars, 1970, Paris.

17 - J.P. BENZECRI (et collaborateurs)
L'analyse des données.
Dunod, 1973.

18 - J.B. BERTHELIN
Le traitement du personnage dans un système de compréhension automatique du récit.
Thèse de 3e cycle, Paris VI.

19 - J.B. BERTHELIN
Communication personnelle

20 - J.L. BINOT, M. GRAITSON, Ph. LEMAIRE, D. RIBBENS
Automatic processing of written french language.
COLING 80 - Tokyo

21 - J.P. BINQUET
Orthographication d'une chaîne phonétique.
Thèse de 3ème cycle - PARIS XI (à paraître)

22 - O. BLOCH et W. VON WARTBURG
Dictionnaire étymologique de la langue française
Presse Universitaire de France (1960).

23 - BOBROW D.G.
Natural language input for a computer problem solving system
in M. Minsky (Ed) "Semantic information processing" Cambridge M.I.T. 1968 pp. 133-215.

24 - C. BOITET
Un essai de réponse à quelques questions théoriques et pratiques liées à la traduction automatique. Définition d'un système prototype.
Thèse d'Etat, Avril 1976, Grenoble.

25 - A. BONNET
Analyse de textes au moyen d'une grammaire sémantique et de schémas. Application à la compréhension de résumés médicaux en langage naturel.
Thèse d'Etat, Janvier 1980, Paris VI.

26 - A. BORILLO
Représentation des connaissances : de l'analyse syntaxique de la phrase à son interprétation in IRIA - Journées de travail "La compréhension" Arc et Senans - Mai 1977 - 17 p.

27 - D. BOURCIER
Information et signification en Droit. Expérience d'une explicitation automatique de concepts.
LANGUAGES n°53 (mars 1979).

28 - R.J. BRACHMAN
What's in a concept : structural foundations for semantics networks
in International Journal Man-Machine Studies n°9 - 1977 - pp. 127-152.

29 - H.E. BREKLE
Sémantique traduit de "Semantik. Eine Einfuhrung in die sprachwissenschaftliche Bedeutungslehre" A. Colin, Paris, 1974.

30 - P. CATALA
L'informatique et la rationnalité du droit
Informatica e diritto 1. Anno VII. Aprile 1981.

31 - J. CELLARD
La semaine prochaine et la prochaine fois
Le Monde Dimanche, 30 Novembre 1980.

32 - E. CHARNIAK
Jack and Janet in search of theory of knowledge
in IJCAI - 20-23 Août 1973, Stanford University. California.

33 - J.C. CHEVALIER, C. BLANCHE-BENVENISTE, M. ARRIVE, J. PEYTARD
Grammaire Larousse du Français contemporain
Larousse, Paris, 1964.

34 - N. CHOMSKY
Syntactic Structures
Mouton - La Haye - 1957 - traduit par BRAUDEAU M. :"Structures syntaxiques"
Seuil, Paris, 1969.

35 - N. CHOMSKY
Aspects of the Theory of Syntax
MIT - Cambridge - Mass. 1965 traduit par MILNER J.C. :"Aspects de la théorie syntaxique", Seuil, Paris, 1971.

36 - A. COLMERAUER
Un sous-ensemble intéressant du français
in IRIA "La compréhension" Arc-et-Senans, Mai 1977, 24 pages.

37 - M.O. CORDIER
Commande d'un robot en langage naturel dans un domaine nécessitant des connaissances pragmatiques : les recettes de cuisine.
Thèse de 3ème cycle, Mars 1979, Universtité de Paris-Sud - Orsay.

38 - D. COULON
Les réseaux sémantiques et les modèles continus
Document IRIA, journées de travail "La compréhension" Arc-et-Senans, Mai 1977.

39 - D. COULON, D. KAYSER
Apport des réseaux sémantiques et des modèles continus dans les systèmes documentaires
in Etat des recherches sur les systèmes d'information documentaire Saint-Vallier, Juin 1978.

40 - D. COULON, J. FUSS, F. JAKOB, D. KAYSER et M. MONFILS
Une expérience de "compréhension" de texte en langage naturel
R.A.I.R.O. Informatique/Computer Science (Vol. 13, n°4, 1979, p. 317 à 349)

41 - D. COULON, D. KAYSER
Représentation informatique de connaissances difficiles à formaliser
Rapport de recherche n° 53, Nov. 1979, Université de Paris-Sud, LRI, ORSAY.

42 - D. COULON, D. KAYSER
Construction de modèles, analogie et approximation
Journées de travail sur l'analogie à la MSH, 21 et 22 Février 1980, Paris.

43 - D. COULON, D. KAYSER
Un système de raisonnement à profondeur variable
Congrès AFCET - Informatique, Nov. 1980.

44 - J. COURTIN
Algorithmes pour le traitement interactif des langues naturelles
Thèse d'Etat, Oct. 1977, Grenoble.

45 - J.M. DAVID
Simulation de l'activité d'un documentaliste.
Congrès AFCET - Informatique, Nov. 1980.

46 - F. DEBILI
Traitements syntaxiques utilisant des matrices de précédence fréquentielles construites automatiquement par apprentissage
Thèse docteur-ingénieur, Paris VII, Septembre 1977.

47 - F. DEBILI
Une méthode de construction automatique de familles de mots
Note C.E.A. - N-2095 - Juillet 1979 - pp. 69.
(International Study Congress on Logic, Informatics, Law - FLORENCE, Avril 1981, pp. 16).

48 - F. DEBILI, C.FLUHR
Similitude entre documents. D'une évaluation pondérée à l'utilisation de relations lexicales sémantiques.
Journées sur l'analogie à la MSH à Paris, 21 et 22 Février 1980.

49 - F. DEBILI
About automatic building of lexicons as a medium for large scale free-text analysis.
Colloque Franco-Suédois Automatique - Informatique - Télématique, Nov. 1980 STOCKHOLM.

50 - F. DEBILI
Représentation des données textuelles obtenues par une analyse syntaxico - sémantique à apprentissage
Colloque sur les bases de données - TUNIS - Avril 1981.

51 - J. DUBOIS
Grammaire structurale du français.
Larousse, Paris, 1965.

52 - J. DUBOIS
Grammaire structurale du français : le verbe
Larousse, Paris, 1967.

53 - J. DUBOIS et F. DUBOIS-CHARLIER
Eléments de linguistique française : Syntaxe
Larousse, Paris, 1970.

54 - J. DUBOIS, L. GUESPIN, M. GIACOMO, Chr. et J.B. MARCELLESI, J.P. MEVEL
Dictionnaire de linguistique
Larousse, Paris, 1973.

55 - O. DUCROT
Dire et ne pas dire
HERMANN.

56 - S.E. FAHLMAN
A system for representing and using real-world knowledge
Artificial Inteligence Laboratory. M.I.T. Déc. 1977.

57 - G. FAUCONNIER
La coréférence : syntaxe ou sémantique ?
Ed. Le Seuil, Paris, 1974.

58 - B. FILLIATRE
Utilisation d'un thésaurus de liaisons analogiques entre concepts dans un système documentaire.
Information processing & management, Vol. 11, pp. 107-113 Pergamon Press 1975.

59 - C.J. FILLMORE
The case for case
in RACH E. & HARMS R. (eds) "Universals in Linguistic Theory" - Holt, Rinehart & Winston Inc., 1968, pp. 1-90.

60 - C. FLUHR
Algorithme à apprentissage et traitement automatique des langues
Thèse d'Etat, juin 1977, Orsay.

61 - C. FLUHR
SPIRIT : A linguistic and probabilistic information storage and retrieval system.
Septembre 1980 - VARSOVIE.

62 - M. GREVISSE
Le bon usage - Grammaire française - 8ème édition
J. DUCULOT, GEMBLOUX, Hatier - Paris (1964).

63 - F. GROSRICHARD
La paresse des sigles
Le Monde Dimanche, 27 avril 1980.

64 - M. GROSS
Méthodes en syntaxe - Hermann - Paris - 1975.

65 - M. GROSS, A. LENTIN
Notions sur les grammaires formelles
Gauthier-Villars, Paris, 1975.

66 - J.P. HATON
Reconnaissance et compréhension automatique du discours parlé : les approches du problème et les tendances en 1977.
in IRIA - Journées de travail "La compréhension" Arc-et-Senans, Mai 1977.

67 - P.J. HAYES
On semantic nets, frames and associations
in IJCAI, MIT Cambridge, Massachusetts, USA, 22-25 August 1977, pp. 99 107.

68 - Y. HLAL
Méthodes d'apprentissage pour l'analyse morphosyntaxique (expérimentées dans le cas de l'arabe et du français)
Thèse d'Etat, Février 1979, Orsay.

69 - B. HUDAULT
Un exemple de compréhension du langage naturel : Sommaires automatiques d'une jurisprudence rendue en matière de responsabilité de la circulation.
Thèse de 3ème cycle, Juillet 1979, Paris VI.

70 - J.H. JAYEZ
Une approche de la compréhension par machine du langage naturel
Thèse d'Etat, Paris VII, Juin 1979.

71 - J.J. KATZ, J.A. FODOR
The structure of a Semantic Theory
in FODOR J.A., KATZ J.J. "The Structure of Language" Prentice-Hall, 1964, pp. 479-518.

72 - D. KAYSER
Informatique et langage naturel
Cours de D.E.A d'informatique. Université de Paris-Sud, Orsay, 1978-1979.

73 - D. KAYSER
Vers une modélisation du raisonnement "approximatif"
Rapport de recherche n°47, Juillet 1979, Université de Paris-Sud, LRI, Orsay.

74 - D. KAYSER
Communication personnelle.

75 - V. KNAPP
Les thésaurus juridiques.
International Study Congress on LOGICA - INFROMATICA - DIRITTO, 6-10 avril 1981, FLORENCE.

76 - J.L. LAURIERE
Raisonnement par analogie en intelligence artificielle : les systèmes de production.
Journées sur l'analogie à la MSH, Paris, 21-22 Février 1980.

77 - J.F. LE NY
Discussion des différents modèles d'un point de vue psychologique
in IRIA, Journées du travail "La compréhension", Arc-et-Senans, 17-18 Mai 1977.

78 - J.F. LE NY
La sémantique psychologique
PUF, Paris, 1979.

79 - J.S. LIENARD
Les processus de la communication parlée
Masson, Paris, 1977.

80 - J. LYONS
Linguistique générale
Larousse, Paris, 1978.

81 - J. LYONS
Eléments de sémantique
Larousse, Paris, 1978.

82 - B. MAEGAARD, E. SPANG-HANSSEN
La segmentation automatique du français écrit
Document de linguistique quantitative n°35, Dunod.

83 - J.J. MARIANI
Contribution à la reconnaissance de la parole utilisant la notion de spectre différentiel
Thèse docteur-ingénieur, mars 1977, Paris VI.

84 - J.J. MARIANI
L'ordinateur qui parle
Le Monde Dimanche, 24 Août 1980.

85 - J.J.MARIANI
Reconnaissance et synthèse de la parole
Micro-systèmes - Sept. Oct. 1980

86 - G. MAZET et C. BERNARD
Traitement par ordinateur des variables allotaxiques
ERA n°96 (Montpellier, 1976).

87 - L. MEHL
Du catalogue à l'analogue et de la taxinomie.
Journées sur l'analogie à la MSH, Paris, 21 et 22 Février 1980.

88 - I.A. MEL'CHUK
Esquisse d'un modèle linguistique du type "sens-texte"
in Problèmes actuels en psycholinguistique, Paris, Ed. du CNRS, 1974.

89 - M. MINSKY
Editor's introduction in MINSKI M. (ed.)
Semantic Information Processing - M.I.T. - Press - Cambridge - Mass. 1968.

90 - M. MINSKY
A framework for representing knowledge
in WINSTON P. (ed.) "The psychology of computer vision" - New-York : McGraw Hill - 1975.

91 - H. MITTERAND
A.B.C de Grammaire Française
Fernand Nathan, Paris, 1969.

92 - G. MOUNIN
Clefs pour la sémantique
Ed. Seghers, Paris, 1972.

94 - J. PITRAT
Présentation des différents niveaux d'analyse
in IRIA "La compréhension" Arc-et-Senans, Mai 1977.

95 - J. PITRAT
La programmation informatique du langage
La recherche n°93, Octobre 1978.

96 - P. PIAT, P. REBOURS
Principes de recherches des distances sémantiques entre mots à l'aide d'un réseau représentant la synonymie d'une manière transitive.
Rapport de stage de D.E.A, Institut de Programmation, Paris VI, Sept. 1973.

97 - P. PLANTE
DEREDEC : Logiciel pour le traitement linguistique et l'analyse de contenu des textes
Université du Québec à Montréal, Mars 1980.

98 - H. PRADE
Note sur l'usage du raisonnement par analogie en Intelligence Artificielle
Journées sur l'analogie à la MLSH, Paris, 21 et 22 Février 1980.

99 - B. PROUTS
Contribution à la synthèse de la parole à partir du texte ; transcription Graphème-Phonème en temps réel sur microprocesseur.
Thèse de docteur-ingénieur, Paris XI, 1980.

100 - M.R. QUILLIAN
Semantic Memory
in Minsky M. (ed.) "Semantic Information Processing" - M.I.T. Press - Cambridge. Ma. - 1968.

101 - G. SABAH
Contribution à la compréhension effective d'un récit
Thèse d'Etat, Paris VI, 1979.

102 - G. SABAH
Un type d'interprétation des métaphores
Journées sur l'analogie à la MSH, Paris, 21 et 22 Février 1980.

103 - M. SALKOFF
Une grammaire en chaîne du français - Analyse distributionnelle
Dunod, Paris, 1973.

104 - M. SALKOFF
L'analyse de la langue écrite et parlée, et ses applications
Ecole informatique de l'IRIA, 12-16 Avril 1976.

105 - M. SALKOFF
Les principes d'un analyseur syntaxique du français
in Linguisticae Investigationes, Vol. n°1, 1977.

106 - R.C. SCHANK
Identification of conceptualizations underlying natural language
in SCHANK R.C., COLBY M.K. (eds) "Computer models of thought and language" Freeman, San Francisco, 1973, pp. 187-248.

107 - R.F. SIMMONS
Semantic Networks : Their Computation and Use for Understanding English Sentences
in R.C. SCHANK and M.K. COLBY (eds) "Computer models of thought and language", Freeman, San Francisco, 1973.

108 - L. TESNIERE
Eléments de syntaxe structurale
Editions KLINCKSIECK, Paris, 1969.

109 - B. VAUQUOIS
La traduction automatique à Grenoble
Document de linguistique quantitative n°24, Dunod, 1975.

110(a) - B. VAUQUOIS
Description de la structure intermédiaire
Comm. Colloque , Luxembourg, Avril 1978.

110(b) - B. VAUQUOIS
Communication personnelle

111 - R.L. WAGNER, J.PINCHON
Grammaire du français classique et moderne
Hachette, Paris, 1962.

112 - Y. WILKS
An artificial intelligence approach to machine translation
in SCHANK R.C., COLBY M.K. (eds) "Computer models of thought and language" Freeman, San Francisco, 1973.

113 - T. WINOGRAD
A procedural model of language understanding
in SCHANK R.C., COLBY M.K. (eds) "Computer models of thought and language" Freeman, San Francisco, 1973.

114 - W.A. WOODS
What's in a link : Foundations for semantic networks
in BOBROW D.G. &COLLINS A.M. (eds) "Representation and Understanding : Studies in Cognitive Science", pp. 35- 82, New-York Academy Press, 1975.

115 - A.K. ZHOLKOVSKIJ and I.A. MEL'CHUK
Sur la synthèse sémantique
T.A. Informations, Revue Internationale du Traitement Automatique du Langage, n°2.
Ed. KLINCKSIECK, Paris, 1970.

116 - R. ZUBER
Structure présuppositionnelle du langage
Document de linguistique quantitative n°17, Dunod, 1972.